Growth Mechanics: General principles of optimal cellular resource allocation in balanced growth

Hugo Dourado[a,✉]**, Wolfram Liebermeister**[b]**, Oliver Ebenhöh**[c]**, and Martin J. Lercher**[a,✉]

[a]Institute for Computer Science and Department of Biology, Heinrich-Heine Universität, 40221 Düsseldorf, Germany
[b]Université Paris-Saclay, INRAE, MaIAGE, 78350 Jouy-en-Josas, France
[c]Quantitative and Theoretical Biology, Heinrich-Heine Universität, 40221 Düsseldorf, Germany

The physiology of biological cells evolved under physical and chemical constraints such as mass conservation, nonlinear reaction kinetics, and limits on cell density. For unicellular organisms, the fitness that governs this evolution is mainly determined by the balanced cellular growth rate. We previously introduced Growth Balance Analysis (GBA) as a general framework to model such nonlinear systems, and we presented analytical conditions for optimal balanced growth in the special case that the active reactions are known. Here, we develop Growth Mechanics (GM) as a more general, succinct, and powerful analytical description of the growth optimization of GBA models, which we formulate in terms of a minimal number of dimensionless variables. GM uses Karush-Kuhn-Tucker (KKT) conditions in a Lagrangian formalism. It identifies fundamental principles of optimal resource allocation in GBA models of any size and complexity, including the analytical conditions that determine the set of active reactions at optimal growth. We identify from first principles the economic values of biochemical reactions, expressed as marginal changes in cellular growth rate; these economic values can be related to the costs and benefits of proteome allocation into the reactions' catalysts. Our formulation also generalizes the concepts of Metabolic Control Analysis to models of growing cells. GM unifies and extends previous approaches of cellular modeling and analysis, putting forward a program to analyze cellular growth through the stationarity conditions of a Lagrangian function. GM thereby provides a general theoretical toolbox for the study of fundamental mathematical properties of balanced cellular growth.

cell growth | self-replication | optimal resource allocation | growth economy | growth control
Correspondence: *hugo.dourado@hhu.de, martin.lercher@hhu.de*

Introduction

A core feature of microbial cells is self-replication – their ability to build a complete, identical cell exclusively out of the chemical compounds found in the environment. If a population of asynchronously replicating microbial cells grows exponentially in a constant environment, its self-replication can often be assumed to result from balanced growth, a non-equilibrium steady state in which every cellular component accumulates at the same rate in proportion to its total amount (1). For non-interacting microbes in a constant environment, the balanced growth rate is equivalent to fitness (2).

The cellular composition is thus often interpreted as an approximate solution to a problem of optimal allocation, driven by natural selection. Accordingly, theoretical methods estimating the optimal allocation are used as a reference to understand cellular composition in vivo (3–8).

At the whole-cell level, a mechanistic understanding of the quantitative principles that shape cellular balanced growth has been approached predominantly through methods collectively classified as constraint-based modeling (CBM). CBM approaches define a solution space of feasible cellular states (usually defined by reaction fluxes) based on simple, mechanistic constraints. The predominant constraint in CBMs is flux balance, encoded through a linear system of equations that constrain the space of allowed reactions fluxes $\mathbf{v}$ (9, 10),

$$\mathbf{S}\,\mathbf{v} = \mathbf{0}. \tag{1}$$

Here, $\mathbf{v}$ is a vector of reaction fluxes, i.e., reaction rates in units of $[\text{moles}][\text{time}]^{-1}[\text{volume}]^{-1}$. Each row of the stoichiometric matrix $\mathbf{S}$ corresponds to one metabolite, while each column corresponds to a metabolic reaction, with entries listing the corresponding stoichiometric coefficients of substrates (negative integers) and products (positive integers).

Thermodynamics and physiological limits – such as a limited nutrient uptake capacity – are typically approximated through fixed upper and/or lower bounds on the modeled fluxes $\mathbf{v}$ (11). The most widely used CBM approach, flux balance analysis (FBA) (11, 12), obtains plausible physiological states by optimizing some objective function over the feasible flux vectors. Frequently, the objective function is the flux through a hypothetical biomass reaction v_{bio}, which mimics the accumulation of precursors for macromolecules and the consumption of energy for their assembly during growth.

Resource balance analysis (RBA) and metabolism and expression models (ME-models) go beyond FBA by aiming to model metabolism in its most general sense, with the ultimate goal of representing all chemical reactions that occur in a living organism (8, 13). In contrast to FBA, these methods take into account the burden of producing the macromolecules required for catalyzing each flux. They approximate the corresponding kinetic rate laws as linear relations between fluxes and the concentration of their catalysts, ignoring the dependence on reactant concentrations (except for dependencies on extracellular concentrations, which serve as model parameters).

All widely used CBMs (8, 11, 13, 14) are formulated as linear optimization problems, which can be solved efficiently even for genome-scale models with thousands of reactions. Accordingly, they are currently the most powerful tools to predict and understand realistic cellular models. However, by construction, these linear methods cannot capture the potentially complex nonlinear relationship between biochemical reaction fluxes – and hence cellular growth — and the concentrations of reactants involved as substrates and products.

Instead of accounting for nonlinear kinetics, these methods rely on linear, phenomenological assumptions.

Nonlinear CBMs (6, 15–19) account for constraints such as nonlinear kinetic rate laws, linking the concentration of metabolites to reaction fluxes. This link means that the metabolite concentrations are now an output of the model instead of an input. Molenaar et al. (6) introduced "self-replicator" models that maximize the cellular growth rate, with reaction fluxes that are limited by fundamental physiological constraints including mass conservation, nonlinear rate laws, and limited protein density. Importantly, these models are completely self-contained, in the sense that in order to grow and self-replicate, all of a model's individual components have to be produced explicitly by the model itself. Instead of using a phenomenological "biomass reaction", the constrained optimization of growth predicts the detailed cell composition, and all possible trade-offs in resource allocation can be accounted for from first principles.

Similar to RBA and ME models, self-replicator models include a "ribosome" reaction that produces the necessary proteins. The proteins can be classified into three categories: transport proteins in the cell surface, which exchange mass with the environment; enzymes, which catalyze internal metabolic reactions; and the ribosome itself, which catalyzes the internal protein production, and which for simplicity is assumed to be composed only of proteins. The study of models of this type relies on the numerical solution of nonlinear optimizations: while it in principle accommodates models with any number of reactions, actual presented models have small, highly simplified reaction networks (6, 15–19).

We have previously formalized a general framework for nonlinear CBM and provided solutions for important special cases, an approach we termed growth balance analysis (GBA) (4) (Fig. 1). GBA models are based on the self-replicator scheme. Instead of a fixed protein concentration, they consider a fixed combined mass density of their components. Optimal cellular resource allocation, as predicted by GBA models, emerges exclusively from quantitative biochemical and physical principles, including the intrinsic nonlinear nature of the underlying reaction kinetics. In general, the optimization of nonlinear models is a non-convex problem, frequently hampered by the existence of multiple local optima (20). Several studies have explored ad-hoc analytical solutions to convex, minimal nonlinear cell models consisting of up to three cellular reactions. Despite their simplicity, simulations with these schematic models are qualitatively consistent with the experimentally observed behaviour of actual cells (6, 15–19).

Together with the GBA model framework, we introduced an analytical theory to study GBA models of any size in the special case that there are no redundant reactions, in the sense that the function of any reaction in the model cannot be copied by any combination of other reactions (i.e., the reaction matrix has full column rank). For this special case, we derived the necessary analytical conditions for optimal growth. Optimal states of arbitrarily complex models also have to satisfy these conditions, as they must always use non-
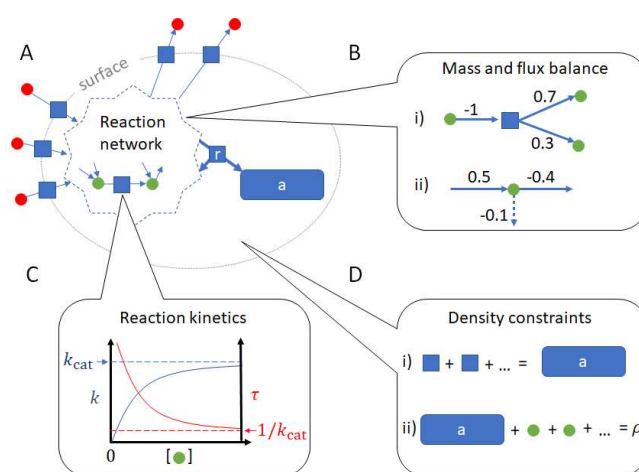

Fig. 1. Constraints in a GBA model. A) In a GBA model, a cell exchanges external reactants (red circles) via transporters (blue squares at the cell surface); converts internal reactants (green circles) via enzymatic reactions (blue squares inside the metabolic network); and produces all proteins catalyzing the reactions (blue rectangle "a") via a ribosome reaction "r". The ribosome reaction consumes and returns metabolites to the metabolic network. In its strict sense, the metabolic network comprises the conversion of small molecules into energy and precursors for macromolecules. A model may also describe metabolism in its more general sense, including other enzymatic reactions such as those for DNA replication and transcription. **B)** All reactions in the model must conserve mass, a concept that comprises (i) mass balance within reactions: one unit of mass consumed (-1) results in one unit of mass produced (+1); and (ii) flux balance of reactant production and consumption, including the dilution by growth of all components (dashed arrow). **C)** Each reaction flux is catalyzed by a specific protein with turnover time τ (or equivalently, turnover rate $k = 1/\tau$). τ is determined by kinetic rate laws and depends on the concentrations of reactants involved in the reaction; $k = 1/\tau$ has a maximal value k_{cat}. **D)** Two basic density constraints govern the cellular interior: (i) the density of proteins "a", and (ii) the total density ρ, which is the sum of all protein and metabolite concentrations.

redundant sets of active reactions (4, 21, 22). The necessary conditions provide a tool to understand important properties of cellular resource allocation at optimality, e.g., the protein allocation into ribosomes at different growth rates (4). On the other hand, the conditions cannot explain themselves why a particular set of non-redundant reactions is used at optimality. In the previously published GBA treatment, this choice has to be an input of the model. This is a major restriction, as real cells activate different sets of reactions depending on the growth conditions, and realistic mathematical models should be able to predict these sets.

Below, we analyze the growth optimization of arbitrary GBA models that may include redundant reactions, generalizing the results in Ref. (4). We provide the necessary analytical conditions for the optimality of each reaction flux, including the criteria for whether a reaction is active at optimality. We interpret these analytical conditions in terms of marginal costs and benefits of reactions with respect to their influence on growth, and quantify how changes in the model parameters and external conditions control the optimal growth rate.

Results

We first present the notation and mathematical definition for growth optimization, including an objective function and constraints. We then reformulate the problem in terms of flux fractions as the only free variables, which greatly simplifies the subsequent analytical study. Finally, we explore the con-

sequences that emerge from the necessary optimality conditions in terms of economics and control theory, and discuss their biological significance. Table 1 lists all symbols used below.

Table 1. Symbols used.

Symbol	Description (units)
A	growth adaptation coefficient
$\mathbf{b}$	biomass fraction vector
$\mathbf{c}$	reactant concentration vector ($[\text{mass}][\text{volume}]^{-1}$)
$\mathbf{C}$	control coefficient matrix
$\mathbf{f}$	flux fraction vector
$\mathbf{E}$	indirect sensitivity matrix ($[\text{time}]$)
$\mathbb{K}$	set of kinetic parameters (various units)
$\mathcal{L}$	Lagrangian ($[\text{time}]^{-1}$)
$\mathbf{M}$	mass fraction matrix
$\mathbf{p}$	protein concentration vector ($[\text{mass}][\text{volume}]^{-1}$)
$\mathbf{S}$	stoichiometric matrix ($[\text{mol}]$)
$\mathbf{v}$	flux vector ($[\text{mass}][\text{volume}]^{-1}[\text{time}]^{-1}$)
$\mathbf{x}$	external concentration vector ($[\text{mass}][\text{volume}]^{-1}$)
γ	vector with column sums of $\mathbf{M}$
Γ	growth control coefficient ($[\text{time}]^{-1}$)
ε	direct sensitivity matrix ($[\text{time}]$)
ϕ	proteome mass fraction vector
λ_j	KKT multiplier for the protein constraint ($[\text{time}]^{-2}$)
λ_m	KKT multiplier for the metabolite constraint ($[\text{time}]^{-1}$)
λ_ρ	KKT multiplier for the density constraint ($[\text{time}]^{-1}$)
μ	growth rate ($[\text{time}]^{-1}$)
$\boldsymbol{\pi}$	parameter vector (various units)
ρ	mass density ($[\text{mass}][\text{volume}]^{-1}$)
$\boldsymbol{\tau}$	turnover time vector ($[\text{time}]$)

Index	Description
a	all proteins
e	enzymatic reactions
i	internal reactants (including m and a)
j	reactions (including s, e, r)
l	reactions (including s, e, r)
m	metabolites
r	ribosome reaction
s	surface reactions (i.e., transport reactions)

Growth Modeling. We define a GBA model as the triple $(\mathbf{M}, \boldsymbol{\tau}, \rho)$. The matrix $\mathbf{M}$ describes the mass fractions of internal reactants consumed and produced by each reaction; $\boldsymbol{\tau}$ is a vector of catalytic turnover times for each reaction; and ρ is the combined mass concentration of all internal components. In the following paragraphs, we provide more detailed descriptions of the model constituents $\mathbf{M}$, $\boldsymbol{\tau}$, and ρ. Here and below, we use the term "reaction" to also encompass transport processes across the cell surface, which are "catalyzed" by transporter proteins or protein complexes.

$\mathbf{M}$ was first introduced in (4). It is constructed from the stoichiometric matrix $\mathbf{S}$ for the total, closed system, i.e., including rows for external reactants. We add a column "r" for the production of protein from precursors by the ribosome, as well as a row "a" corresponding to the total concentration of all proteins, following the procedure introduced by Molenaar et al. (6). We now first convert all entries to masses, by multiplying each row with the corresponding molecular mass. Because of mass balance, each column must then sum to 0. We next normalize each column such that the sum of negative entries equals -1 and the sum of positive entries equals $+1$. Now the entries correspond to the mass fractions of each reactant (rows) going into and out of each reaction (columns), as illustrated for the example in Fig. 1B. Finally, we reduce the normalized matrix to a matrix for an open system, by dropping all rows for reactants external to the modeled cell. For the remaining internal reactants, we will assume a quasi-steady state and thus enforce mass conservation.

As illustrated in Fig. 2, to simplify the notation for the following theoretical development, we partition the columns of $\mathbf{M}$ (indexed together by j) into index sets for reaction types: s for transport processes across the cell surface; e for internal enzymatic reactions; and r for the ribosome reaction, which is the only one that produces protein. We partition the rows of internal reactants (indexed together by i) into indices m for metabolites and a for total protein. We use the term "metabolites" in its more general sense, referring to any molecule in the cell that is not a protein. We distinguish vectors by using boldface, and vector components by using italics with the appropriate index, e.g., $\mathbf{c}$ is the column vector of all internal reactant concentrations, c^i are its components, and we use a lower index to indicate the components c_i of the row vector $\mathbf{c}^\top$.
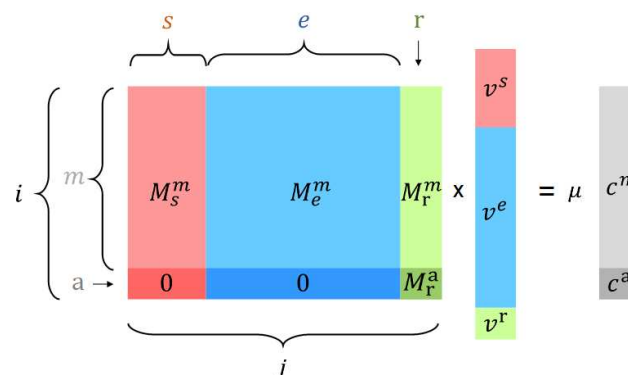

Fig. 2. Schematic overview of the mass conservation constraint $\mathbf{M}\mathbf{v} = \mu\mathbf{c}$, determined by the mass fraction matrix $\mathbf{M}$, the column vector of mass fluxes $\mathbf{v}$, the growth rate μ, and the column vector of internal reactants mass concentrations $\mathbf{c}$. The indices indicate partitions according to the type of reaction (columns of $\mathbf{M}$, $\mathbf{v}$) or reactant (rows of $\mathbf{M}$, $\mathbf{c}$). The index $i = (m, \text{a})$ correspond to rows for internal reactants, comprising rows m for metabolites, and a row "a" for the total mass concentration of all proteins. The indices $j = (s, e, \text{r})$ correspond, respectively, to transport proteins, enzymes, and the ribosome. We also use the index l for all reactions when necessary. Note the row "a" of $\mathbf{M}$ has only one nonzero entry M_r^a, corresponding to the mass fraction of protein produced by the ribosome reaction r. Different colors indicate three different types of reactions: red (transporters), blue (enzymes), green (ribosome); and two types of reactants: light gray (metabolites), dark gray (total protein), resulting in six partitions of $\mathbf{M}$.

The ribosome reaction represents the last step in protein synthesis, and is assumed to be catalyzed by a "ribosome" consisting entirely of protein. We here ignore the RNA components of the ribosome for simplicity, but it is possible to extend the modeled ribosome to a more realistic RNA-protein complex. In addition, the enzymatic reactions (e) could be extended so that they include details of protein translation

that occur before the last, "ribosome" step (r) (5). Note that nonlinear genome-scale GBA models can be created from existing linear genome-scale models, by extending their stoichiometric matrix $\mathbf{S}$ with the addition of a ribosome reaction, normalizing it to $\mathbf{M}$ with the molecular masses, and adding the kinetic rate laws $\boldsymbol{\tau}$ and density ρ (see below).

We assume that every reaction represented by a column j in $\mathbf{M}$ is catalyzed by a protein or protein complex with concentration p^j – specifically, a transporter (s), an enzyme (e), or the ribosome (r). The corresponding flux v^j is assumed to be proportional to p^j, expressed as $v^j = p^j/\tau^j(\mathbf{c}, \mathbf{x})$. Thus, τ^j is defined as the inverse of the usual metabolite-dependent factor in kinetic rate laws; SI text "Rate laws and kinetic parameters" provides a basic discussion of rate laws and the necessary kinetic parameters. $\mathbf{c} = (c^m, c^a)^\top$ is the vector of internal reactant concentrations, comprising all metabolite concentrations c^m as well as c^a, the combined mass concentration of all proteins. Hence, each τ^j may depend not only on the concentrations of the substrates and products of the corresponding reaction, but also on inhibitors and regulatory metabolites not involved in the turnover itself. The transport processes s are the only reactions whose rate laws may depend on the external concentrations $\mathbf{x}$.

Note that in accordance with the normalization of $\mathbf{M}$, all concentrations of proteins p^j and reactants c^i throughout this work are in units of $[\text{mass} \times \text{volume}^{-1}]$. Fluxes $([\text{mass} \times \text{volume}^{-1} \times \text{time}^{-1}])$ and the kinetic parameters must then also be expressed in mass units, e.g., Michaelis constants K_m in $[\text{mass} \times \text{volume}^{-1}]$ and turnover numbers k_cat in product mass per protein mass per time, resulting in $[\text{time}^{-1}]$.

ρ, the final constituent of GBA models, is the sum of all internal concentrations. We assume ρ to be constant, which is consistent with experimental data on *E. coli* across growth conditions and even across the cell cycle (23–25). The mass balances exploited for the normalization of $\mathbf{M}$ mean that all reactants involved in reactions must be accounted for in the model and hence be included in the value of ρ; e.g., in a realistic model water is a reactant in many reactions, so ρ corresponds in this case to the total cellular density (or buoyant density). Simplified models may instead include only dry mass components, so that both $\mathbf{M}$ and ρ consider only these. The equations below will be simplified by adopting the Einstein summation convention: a repeated lower and upper index denotes a summation over this index (often indicating a matrix multiplication), e.g., $M^i_j v^j := \sum_j M_{ij} v_j = [\mathbf{Mv}]_i$. We also denote element-wise multiplication with repeated upper (or repeated lower) indices.

Mass conservation implies that in the mass fraction matrix $\mathbf{M}$, each column sum $\gamma_j := \sum_i M^i_j$ is zero if it involves only the consumption and production of internal reactants (indices e, r). In contrast, transport reactions (with indices s), which bring mass into and out of the modeled system, do not conserve mass, resulting in the equations

$$\gamma_r = 0$$
$$\gamma_e = 0 \qquad \qquad \textbf{(2)}$$
$$\gamma_s \neq 0 \quad .$$

The property (2) guarantees mass conservation within reactions, an information that is not always fully encoded in the stoichiometric matrix $\mathbf{S}$ (see SI text "Mass balance and the stoichiometric matrix $\mathbf{S}$"). While external reactants have no corresponding rows in $\mathbf{M}$, their concentrations $\mathbf{x}$ may influence the turnover times τ^s of transporters. We present examples of GBA models and R code for their numerical optimization in Dataset S1 (see SI text "Examples of GBA models and R code for numerical optimization").

We are interested in the cellular physiology, defined through the concentration vectors $\mathbf{c}, \mathbf{p}$ and the vector of reaction fluxes $\mathbf{v}$, at balanced growth. For a given model specified by $(\mathbf{M}, \boldsymbol{\tau}, \rho)$ and a given environment specified by $\mathbf{x}$, balanced growth at the instantaneous rate μ is specified by the following constraints:

$$M^i_j v^j = \mu c^i \qquad \text{(mass conserv. in bal. growth)} \quad \textbf{(3)}$$
$$v^j = p^j/\tau^j(\mathbf{c}, \mathbf{x}) \quad \text{(reaction kinetics)} \quad \textbf{(4)}$$
$$\sum_j p^j = c^a \qquad \text{(protein density)} \quad \textbf{(5)}$$
$$\sum_i c^i = \rho \qquad \text{(cellular density)} \quad \textbf{(6)}$$
$$c^m \geq 0 \qquad \text{(non-negative met. conc.)} \quad \textbf{(7)}$$
$$p^j \geq 0 \qquad \text{(non-negative protein conc.)} \quad \textbf{(8)}$$

The model's balanced growth property is captured by the right hand side of Eq. 3. We assume that the growth rate is always positive, $\mu > 0$. Thus, for internal nodes with non-zero concentration $(c^i \neq 0)$, there is a necessary mass flow to offset the dilution through the associated volume growth at rate μ. Note that the total protein concentration c^a in Eq. 5 is not fixed, but is constrained by a row "a" in Eq. 3 that specifies mass conservation in balanced growth.

Below, we will be interested in optimal balanced growth, defined as growth at the maximal possible rate μ given the above constraints. From Eq. 3-5, we see that the variables $(\mathbf{c}, \mathbf{p}, \mathbf{v})$ are highly interdependent. The above formulation does not lend itself to expressing μ as an explicit function of these variables, which makes it not ideal for numerical or analytical studies. If one can find a mathematically equivalent formulation based on fewer, independent variables, then this would facilitate the use of the KKT conditions, analogous to how generalized coordinates facilitate the solution of problems in Lagrangian mechanics (26). Thus, we next focus on a corresponding reformulation of the optimization problem. This formulation will apply to all balanced growth states, and only later we will use it to examine optimal balanced growth states.

A Reformulation in Terms of Flux Fractions f. Our guiding thought below is that there can be a correspondence between cell states at different growth rates, which can be expressed in the form of scaling relations. These scaling relations extend the mass fraction scaling of $\mathbf{M}$ to fluxes and concentrations. Specifically, we define *biomass fractions*

$$\mathbf{b} := \frac{\mathbf{c}}{\rho} \quad \text{(adimensional)} \quad \textbf{(9)}$$

(equivalent to $\mathbf{c} = \rho\,\mathbf{b}$, since $\rho > 0$), which express concentrations as fractions of the total cellular density; and we define *flux fractions*

$$\mathbf{f} := \frac{\mathbf{v}}{\mu\rho} \quad \text{(adimensional)} \tag{10}$$

(equivalent to $\mathbf{v} = \mu\rho\,\mathbf{f}$, since we assume $\mu\rho > 0$), which express fluxes as fractions of the net mass uptake – i.e., the net growth – of the cell, $\mu\rho$. Thus, each flux fraction f^j describes the activity of reaction j relative to the total cellular mass production. Importantly, Eq. 3 describing mass conservation in balanced growth does not depend explicitly on μ anymore when written in terms of $\mathbf{f}$ and $\mathbf{b}$:

$$\mathbf{M}\mathbf{f} = \mathbf{b} \quad . \tag{11}$$

This equation also implies that the mass fractions $\mathbf{b}$ are uniquely determined by the flux fractions $\mathbf{f}$, independently of μ. Conveniently, this unique dependence also means that we can express the turnover times as functions of only $\mathbf{f}$ and the fixed parameters ρ, $\mathbf{M}$, and $\mathbf{x}$:

$$\boldsymbol{\tau} = \boldsymbol{\tau}(\mathbf{c},\mathbf{x}) = \boldsymbol{\tau}(\rho\mathbf{M}\mathbf{f},\mathbf{x}) \quad . \tag{12}$$

In the following discussion, we mostly focus on the dependence of $\boldsymbol{\tau}$ on $\mathbf{f}$, and for simplicity of notation we do not state the dependence of $\boldsymbol{\tau}$ on the fixed parameters $(\rho,\mathbf{M},\mathbf{x})$ explicitly. Importantly, $\boldsymbol{\tau}$ does not depend explicitly on μ, which otherwise would cause a recursion problem when further expressing the growth rate μ in terms of only $\mathbf{f}$ and $\boldsymbol{\tau}(\mathbf{f})$, as we will see below. The resulting dimensionality reduction of the solution space not only simplifies the analytical considerations below, but also potentially facilitate numerical optimizations (27).

From Eqs. 4 and 10, we obtain the expression for protein concentrations in terms of $\mathbf{f}$, μ, and ρ,

$$p^j = \mu\,\rho\,f^j\,\tau^j(\mathbf{f}) \quad . \tag{13}$$

The combined mass fraction of all proteins in the cell, b^a, is the sum of all p^j in the last equation, divided by ρ:

$$b^a = \mu\,f_j\,\tau^j(\mathbf{f}) \quad , \tag{14}$$

where we indicate summation by combining the lower index in f_j with the upper index in τ^j. Thus, we can calculate the total protein mass fraction during balanced growth from μ and $\mathbf{f}$, based only on reaction kinetics. However, through Eq. 11, the same total protein mass fraction is also related to $\mathbf{f}$ through the corresponding row "a" in $\mathbf{M}$:

$$b^a = M_j^a\,f^j = M_r^a\,f^r \quad , \tag{15}$$

where the second equality reflects our assumption that the "ribosome" reaction r is the only one producing proteins, so that $M_j^a = 0$ for $j \neq r$. Equating the right hand sides of the previous two equations and solving for μ (with $b^a \neq 0 \Rightarrow f_j\,\tau^j(\mathbf{f}) \neq 0$), we get the *growth function*

$$\mu(\mathbf{f}) = \frac{M_r^a\,f^r}{f_j\,\tau^j(\mathbf{f})} \quad . \tag{16}$$

Thus, the growth rate becomes an explicit function of only the flux fractions $\mathbf{f}$. μ still depends on the fixed parameters ρ, $\mathbf{M}$, and $\mathbf{x}$ through the functions $\boldsymbol{\tau} = \boldsymbol{\tau}(\rho\mathbf{M}\mathbf{f},\mathbf{x})$. Note that if fluxes $\mathbf{v}$ were used instead of the flux fractions $\mathbf{f}$, then $\boldsymbol{\tau}(\mathbf{c},\mathbf{x}) = \boldsymbol{\tau}(\mathbf{M}\mathbf{v}/\mu,\mathbf{x})$, which would cause a recursion issue when defining the growth rate as a function of $\mathbf{v}$ and $\boldsymbol{\tau}$ following the same procedure. In that case, one is forced to account for $\mathbf{c}$ as separate variables, thereby increasing the dimensionality of the problem. The same recursion issue occurs when formulating the problem in terms of protein concentrations.

From now on, we will consider $\mathbf{b}$ (Eq. 11) and $\boldsymbol{\tau}$ (Eq. 12) as functions of $\mathbf{f}$, and treat $\mathbf{f}$ as the only free variables. After writing the growth rate μ as a function of $\mathbf{f}$, we now do the same thing for our remaining constraints, so now we have much fewer variables and constraints.

In the scaled variables, the density constraint (Eq. 6) is reduced to

$$\sum_i b^i = 1 \quad . \tag{17}$$

Using the balanced growth Eq. 11, we can rewrite this constraint in terms of flux fractions $\mathbf{f}$. We see that in balanced growth, the density constraint (Eq. 17) is equivalent to a flux balance on the cell surface,

$$1 = \gamma_j\,f^j = \gamma_s\,f^s \quad . \tag{18}$$

The second equality comes from Eq. 2: only the columns s sum up to non-zero values γ_s, so only transport fluxes f^s are limited by this constraint. The nature of this constraint as a global mass balance becomes more evident if we multiply the whole expression by $\mu\rho$: the net mass uptake $\gamma_s\,v^s$ going through the cell surface must equal the rate of biomass production $\mu\rho$.

Any solution to the growth function (Eq. 16) automatically respects internal mass conservation, protein density and the kinetic constraints: for any given vector $\mathbf{f}$, $\mu(\mathbf{f})$ returns the unique growth rate satisfying these constraints (which also depend on ρ through $\boldsymbol{\tau} = \boldsymbol{\tau}(\rho\mathbf{M}\mathbf{f})$). The flux balance at the cell surface is enforced separately by Eq. 18 on transporters, making these fundamentally different from enzymatic and ribosome reactions. In particular, for a model with only one transporter s, Eq. 18 already determines the scaled uptake rate $f_s = (\gamma_s)^{-1}$. With two transport fluxes, one flux is uniquely determined by the other; a simple example would be a model that only has transporters for glucose uptake and CO_2 excretion (see example model "C" in SI text "Examples of GBA models and R code for numerical optimization"). More generally, Eq. 18 can be used to uniquely determine one transport flux fraction in terms of the others, reducing the number of variables by one. For clarity of presentation, however, we will keep Eq. 18 as a separate constraint and not eliminate any variable, until the introduction of growth control coefficients in the corresponding section.

Finally, writing the non-negativity constraints on proteins and metabolite concentrations in terms of $\mathbf{f}$ results in

$$f^j\,\tau^j(\mathbf{f}) = \frac{p^j}{\mu\rho} \geq 0 \quad , \tag{19}$$

$$M_j^m f^j = b^m = \frac{c^m}{\rho} \geq 0 \quad . \tag{20}$$

The non-negativity of the total protein concentration does not have to be imposed explicitly, as it already follows from the non-negativity of individual proteins. We are now in the position to provide a concise formulation of growth rate optimization in terms of flux fractions $\mathbf{f}$. Combining Eqs. 16, 18, 19, 20, the optimal growth problem for a given environment $\mathbf{x}$ becomes

$$\underset{\mathbf{f}}{\text{maximize}} \quad \mu(\mathbf{f}) = \frac{M_r^a f^r}{f_j \tau^j(\mathbf{f})}$$

subject to:

$$\begin{aligned} \gamma_s f^s &= 1 \\ f^j \tau^j(\mathbf{f}) &\geq 0 \\ M_j^m f^j &\geq 0 \quad , \end{aligned} \tag{21}$$

where the turnover times τ^j are functions that depend on $\mathbf{f}$ and on the parameters $\rho, \mathbf{M}, \mathbf{x}$, $\tau^j = \tau^j(\rho \mathbf{M} \mathbf{f}, \mathbf{x})$. If the rate laws have a general functional form, these functions will be parameterised by the set kinetic parameters $\mathbb{K}$. After solving this optimization problem, all original cellular variables (unscaled fluxes as well as unscaled metabolite and protein concentrations) can be easily reconstructed from $\mathbf{f}$. In the following, we will refer to $\boldsymbol{\pi}$ as the vector of parameters that define the optimization problem, which includes $\mathbf{M}, \rho$, and $\mathbf{x}$, as well as the elements of $\mathbb{K}$. The parameters in $\boldsymbol{\pi}$ are considered fixed until the section "Growth Control and Adaptation", where we study the sensitivity of optimal growth to marginal changes in the components of $\boldsymbol{\pi}$.

Growth Analysis. Next, we utilize the problem reformulation to derive general necessary conditions of cellular systems at optimal balanced growth, independent of model specifics. First, for each reaction, we will derive explicit expressions for shadow prices in the optimal state. We then derive equations for the state variables $\mathbf{f}$ themselves, which must hold in any optimal state. This development constitutes a generalization of our previous analytical approach to GBA (4), which was restricted to models with matrices $\mathbf{M}$ of full column rank. The latter condition is not generally satisfied by realistic cellular models, as many cellular biochemical reactions are structurally redundant, i.e., their columns in $\mathbf{M}$ are linearly dependent on other columns. Optimal growth states always have non-redundant active reactions (i.e., the *active* $\mathbf{M}$ has full column rank) (4, 21, 22), but this optimal set of active reactions is generally not known *a priori*. In contrast, the following analysis in terms of flux fractions is valid for any $\mathbf{M}$ of arbitrary size and rank.

For the following, we emphasize that the state of our system is completely determined by scaled fluxes f^j, which serve as independent variables. All other variables are fully dependent on them: the unscaled fluxes $\mathbf{v}$, the scaled and unscaled concentrations $\mathbf{b}, \mathbf{c}$, and $\mathbf{p}$, the reaction times $\boldsymbol{\tau}$, and the growth rate μ.

All following analyses benefit from knowing the system's sensitivity to small changes of each of the independent variables f^j. The partial derivatives of the system's properties

$\mathbf{c}(\mathbf{f}), \mathbf{v}(\mathbf{f}), \mathbf{b}(\mathbf{f}), \boldsymbol{\tau}(\mathbf{f}), \mathbf{p}(\mathbf{f})$, and $\mu(\mathbf{f})$ with respect to each f^j provide explicit expressions for sensitivity coefficients similar to the ones introduced in Metabolic Value Theory (28, 29), based on the original concepts of Metabolic Control Analysis (MCA) (9, 10). A unique feature of the present treatment arises from the system of equations in Eq. 11, which determines the linear dependence of biomass fractions $\mathbf{b}$ on $\mathbf{f}$, so that the partial derivative of b^i with respect to f^j is given simply as

$$\frac{\partial b^i}{\partial f^j} = M_j^i \quad . \tag{22}$$

Via the chain rule of differentiation, this expression also determines the partial derivatives with respect to f^j for any functions of b^i. A case of particular interest in the following discussions is the vector of turnover times $\boldsymbol{\tau} = \boldsymbol{\tau}(\mathbf{c}, \mathbf{x}) = \boldsymbol{\tau}(\rho \mathbf{b}, \mathbf{x}) = \boldsymbol{\tau}(\rho \mathbf{M} \mathbf{f}, \mathbf{x})$. We first define the (direct) *time elasticities* (*elasticities* in short), the sensitivity of each turnover time $\tau^l(\mathbf{c}, \mathbf{x}) = \tau^l(\rho \mathbf{b}, \mathbf{x})$ with respect to each biomass fraction b^i, as

$$\varepsilon_i^l := \frac{\partial \tau^l}{\partial b^i} = \frac{\partial \tau^l}{\partial c^i} \frac{\partial c^i}{\partial b^i} = \frac{\partial \tau^l}{\partial c^i} \rho \quad , \tag{23}$$

where we used the chain rule of differentiation in the first equality and Eq. 9 in the second. We then use the direct elasticities ε_i^l to express the sensitivity of τ^l to a change in a flux fraction f^j, defined as the *indirect time elasticity* matrix $\mathbf{E}$ (or *indirect elasticity* in short), with entries

$$E_j^l := \frac{\partial \tau^l}{\partial f^j} = \frac{\partial \tau^l}{\partial c^i} \frac{\partial c^i}{\partial b^i} \frac{\partial b^i}{\partial f^j} = \varepsilon_i^l M_j^i \quad , \tag{24}$$

where we sum over i and used Eq. 22 in the last equality. In the following discussion, we assume that the kinetic rate laws do not depend on the total protein concentration c^a, meaning $\varepsilon_a^l = \partial \tau^l / \partial c^a = 0$ for all reactions l. That would be different if, for example, one accounts for the macromolecular crowding effects via kinetic rate laws (30). The indirect elasticities $\mathbf{E}$ and direct elasticities $\boldsymbol{\varepsilon}$ share some resemblance with the Jacobian and elasticity matrices defined in Metabolic Value Theory and MCA, although we do not intend to explore the exact relationships in this work. For an example of direct and indirect elasticities, where τ follows a simple Michaelis-Menten rate law, see SI text "Rate laws and kinetic parameters".

In the remainder of this paper, we will explore three complementary types of analyses of GBA systems. First, in the growth optimality section we will state the analytical conditions necessary for an optimal state $\mathbf{f}^*$. Second, in the growth economy section we will calculate the sensitivity of a (not necessarily optimal) growth rate μ to small changes in each $\mathbf{f}$, which we interpret in economic terms as marginal values of reactions. Third, in the growth control and adaptation section we will estimate the sensitivity of the optimal growth rate μ^* to small changes in the previously fixed parameters $\boldsymbol{\pi}$. In each of these analyses, the sensitivity measures captured by $\mathbf{E}$ will appear naturally in the results.

Growth Optimality. We next calculate the necessary analytical conditions for the optimal growth state. This calculation extends our previous analytical approach, which was restricted to GBA models with matrices $\mathbf{M}$ of full column rank (4), to general GBA models with arbitrary matrices $\mathbf{M}$, facilitated by the reformulation of the GBA problem in terms of flux fractions $\mathbf{f}$. We approach this problem by studying the Karush-Kuhn-Tucker (KKT) conditions (31, 32), which generalize the method of Lagrange multipliers by also accounting for inequality constraints, here present due to the non-negativity of concentrations. To simplify the presentation in this section, we here account explicitly only for the non-negativity of protein concentrations, but not for the non-negativity of metabolite concentrations. Under the reasonable assumption that metabolites with zero concentration do not participate in any active reactions, the resulting necessary conditions are also necessary when accounting for this latter constraint; the full calculations can be found in the Materials and Methods section.

We define the Lagrangian $\mathcal{L}(\mathbf{f}, \boldsymbol{\lambda})$ for a given GBA model $(\mathbf{M}, \boldsymbol{\tau}, \rho)$ and external concentrations $\mathbf{x}$ as

$$\mathcal{L}(\mathbf{f}, \boldsymbol{\lambda}) := \mu(\mathbf{f}) + \lambda_\rho \left(\gamma_s f^s - 1\right) + \lambda_l f^l \tau^l(\mathbf{f}) \quad, \tag{25}$$

where we sum over l in the last term. The *KKT multipliers* $\boldsymbol{\lambda}$ are auxiliary variables used to find the optimal state, but also encode important economic and control information about the system at optimality, as we will see later. λ_ρ relates to the density constraint connected to $\mathbf{f}$ via the flux balance at the surface (Eq. 18); the λ_l relate to the non-negativity of proteins (Eq. 19). The necessary KKT conditions include the primal feasibility conditions given by these two equations and

$$\partial_j \mathcal{L} = 0 \qquad \text{(stationarity)} \tag{26}$$

$$\lambda_j f_j \tau_j = 0 \qquad \text{(complementary slackness)} \quad, \tag{27}$$

where $\partial_j := \partial/\partial f^j$ indicates the partial derivative with respect to f^j.

The stationarity conditions can be solved for the corresponding optimal multipliers λ_j, resulting in

$$\lambda_j = -\left(\partial_j \mu + \lambda_\rho \gamma_j\right)/\tau_j \quad, \tag{28}$$

where λ_ρ is the optimal value for the density multiplier. After an element-wise multiplication of both sides of Eq. 28 with $f_j \tau_j$, we can use the complementary slackness ($\lambda_j \tau_j f_j = 0$) to get

$$\left(\partial_j \mu\right) f_j + \lambda_\rho \gamma_j f_j = 0 \quad. \tag{29}$$

Now summing the last equation over all j and using the primal feasibility (Eq. 18) results in

$$\lambda_\rho = -\left(\partial_j \mu\right) f^j \quad, \tag{30}$$

where we have a summation over j.

From Eq. 16, the partial derivative of μ with respect to f^j is

$$\partial_j \mu = \frac{\mu}{b^{\mathrm{a}}} \left(M_j^{\mathrm{a}} - \mu \tau_j - \mu f_l E_j^l\right) \quad, \tag{31}$$

where we used the identity $b^{\mathrm{a}} = M_{\mathrm{r}}^{\mathrm{a}} f^{\mathrm{r}}$ from Eq. 15. Substituting these derivatives into Eq. 30) results in an expression for the optimal value of the density constraint multiplier λ_ρ in terms of $\mathbf{f}$ only,

$$\lambda_\rho = \frac{\mu^2}{b^{\mathrm{a}}} f_l E_j^l f^j \tag{32}$$

(see Materials and Methods for the detailed calculations). Note that in the last equation, j and l do not refer to specific reactions but instead indicate summations. When we further consider that only transporters have a nonzero column sum (Eq. 2), we get an equivalent expression that highlights the particular dependence of λ_ρ on the reactions directly connected to the transport reactions (see SI text "The dependence of λ_ρ on transporters").

Combining Eq. 28, 31, 32, we can now express each multiplier λ_j explicitly in terms of the flux fractions $\mathbf{f}$ at optimality, resulting in slightly different expressions for ribosomal, enzymatic, and transport reactions:

$$\lambda_{\mathrm{r}} = -\frac{\mu}{b^{\mathrm{a}}} \frac{1}{\tau_{\mathrm{r}}} \left(M_{\mathrm{r}}^{\mathrm{a}} - \mu \tau_{\mathrm{r}} - \mu f_l E_{\mathrm{r}}^l\right) \tag{33}$$

$$\lambda_e = \frac{\mu}{b^{\mathrm{a}}} \frac{1}{\tau_e} \left(\mu \tau_e + \mu f_l E_e^l\right) \tag{34}$$

$$\lambda_s = \frac{\mu}{b^{\mathrm{a}}} \frac{1}{\tau_s} \left(\mu \tau_s + \mu f_l E_s^l - \mu f_l E_j^l f^j \gamma_s\right) \quad. \tag{35}$$

By inserting these expressions into the complementary slackness conditions (Eq. 27), we can now solve for $\mathbf{f}$, which results in the three *balance equations* for ribosomal, enzymatic, and transport reactions:

$$\left(M_{\mathrm{r}}^{\mathrm{a}} - \mu \tau_{\mathrm{r}} - \mu f_l E_{\mathrm{r}}^l\right) f_{\mathrm{r}} = 0 \tag{36}$$

$$\left(\mu \tau_e + \mu f_l E_e^l\right) f_e = 0 \tag{37}$$

$$\left(\mu \tau_s + \mu f_l E_s^l - \mu f_l E_j^l f^j \gamma_s\right) f_s = 0 \quad, \tag{38}$$

where we simplified the expressions by exploiting that $\mu, b^{\mathrm{a}}, \tau_j \neq 0$. The balance equations are necessary equality conditions for the optimal growth state $\mathbf{f}^*$ of the optimization in Eq. 21, and generalize Eq. 10 in (4). Note that in all cases, a reaction with nonzero flux f^j requires that the corresponding term in parentheses (i.e., the corresponding λ_j) is equal to zero. Conversely, if the term in parentheses is different from zero ($\lambda_j \neq 0$), then the reaction cannot carry flux at optimality ($f^j = 0$). In particular, the ribosome evidently needs to be active for balanced cellular growth, so $\lambda_{\mathrm{r}} = 0$ must always hold in optimal states. The KKT multipliers λ_j can be understood as *shadow prices* of the scaled protein concentrations $\tau^j f^j = p^j/(\mu\rho)$ (protein concentration per biomass production $\mu\rho$). These shadow prices must be zero for active reactions at optimality, as otherwise growth rate could be increased by changing the corresponding protein concentration.

We may also express the balance equation for each reaction j in the usual, unscaled variables $\mathbf{p}$ (protein concentrations), $\mathbf{v}$ (fluxes), and $\mathbf{c}$ (reactant concentrations, including metabolites and total protein), by using Eqs. 4,9,10,11 (see Materials and Methods)

$$M_j^{\mathrm{a}} v_j - \mu p_j - \frac{p_l}{\tau_l}\frac{\partial \tau^l}{\partial c^i}M_j^i v_j + \frac{p_l}{\tau_l}\frac{\partial \tau^l}{\partial c^i}\frac{c^i}{\rho}\gamma_j v_j = 0 \quad , \quad \textbf{(39)}$$

where we sum over indices l and i. We now continue our analysis in terms of the flux fractions $\mathbf{f}$, since these are the variables of the optimization problem (Eq. 21). However, we keep in mind that the same change of variables to $\mathbf{p}, \mathbf{v}, \mathbf{c}$ is possible in all the following equations, as done for Eq. 39.

Growth Economy. As growth rate is closely related to fitness (2), it makes sense to view growth rate as the primary value of the cellular economy. In this subsection, we will thus explore the economy of balanced cellular growth, by asking how a small change in the state variables f^j affects the growth rate μ of any optimal or non-optimal state. Below, we will see that the necessary conditions of optimal growth specify that the marginal costs and benefits of each flux must be perfectly balanced.

We define the *marginal value* of flux j as the partial derivative $\partial_j\mu$, which quantifies the marginal gain in growth rate resulting from a small increase in f^j. From Eq. 31, we see that the marginal value can be expressed naturally as a multiple of the growth rate per mass fraction of protein in biomass, μ/b^{a}. As we will see next, the corresponding adimensional factor — the term in parentheses in Eq. 31 – quantifies different types of costs (when negative) and benefits (when positive) of reaction j in terms of its influence on protein allocation,

$$\frac{b^{\mathrm{a}}}{\mu}\partial_j\mu = M_j^{\mathrm{a}} - \mu\tau_j - \mu f_l E_j^l \quad . \qquad \textbf{(40)}$$

The first summand quantifies how a marginal increase in f^j increases the total protein fraction in the cell density $b^{\mathrm{a}} = c^{\mathrm{a}}/\rho$ (see Eq. 5),

$$M_j^{\mathrm{a}} = \frac{\partial b^{\mathrm{a}}}{\partial f^j} \quad . \qquad \textbf{(41)}$$

We name this contribution to the normalized marginal value the *marginal protein production*. As we assume that the ribosome reaction is the only reaction that consumes or produces protein, this reaction ($j = \mathrm{r}$) is the only one with a nonzero (and positive) marginal production benefit.

To interpret the remaining summands, we remember that an individual protein's mass fraction in the cellular density can be expressed as $p^l/\rho = \tau^l v^l/\rho = \mu\tau^l f^l$. The last two terms in Eq. 40 quantify the combined decrease of individual protein fractions in cellular density (p^l/ρ) caused by a marginal increase in f^j at fixed μ,

$$-\mu\tau_j - \mu f_l E_j^l = -\left(\frac{\partial(\mu f_l \tau^l)}{\partial f^j}\right)_\mu = -\sum_l\left(\frac{\partial(p^l/\rho)}{\partial f^j}\right)_\mu \quad .$$
$$\textbf{(42)}$$

Here, the first summand quantifies the change in p^l/ρ at fixed turnover times, which is evidently non-zero only for

the enzyme catalyzing the perturbed flux j itself. We name this term, $-\mu\tau_j$, the *marginal (protein) investment* into j. The final summand quantifies the local change of the individual protein concentrations that must occur to compensate the changes in the turnover times (quantified by the indirect elasticity $\mathbf{E}$), themselves caused by changes in metabolite concentrations forced due to flux balance. We name it the *marginal (protein) opportunity* of j, as it is related to opportunity costs and benefits in economics. For the typical case of reactions running in the forward direction ($f^j > 0$), τ^j is positive, and thus the marginal investment into j is negative, representing a cost. If all fluxes are non-negative, beneficial decreases in turnover times correspond to negative E, resulting in positive marginal opportunity (i.e., marginal opportunity benefit).

We can now summarize our insights about cellular economy, in particular about changes in the growth rate μ in response to changes in a flux f^j. The first and second terms in Eq. 40 are simple, direct consequences of the flux change: the marginal production benefit, an increase in protein production if f^j is the ribosome flux; and the marginal investment, an increase in the protein concentration required to sustain an increased f^j. The third term in Eq. 40, the marginal opportunity, is more interesting, though equally easy to understand. As a simple consequence of mass conservation (Eq. 11), a change in f^j while keeping all other fluxes fixed must result in changes in the concentrations of all reactants consumed or produced in the corresponding reaction. These concentration changes modify the turnover times $\tau_l(\mathbf{c})$ of all reactions l whose kinetics depend on them, either because they are directly connected to those reactants or because they act as inhibitors or activators; see Fig. 3 for an example. Keeping the corresponding fluxes f^l constant requires matching changes in the concentrations p_l of the catalyzing proteins (Eq. 4). This total amount of "protein saved" due to a change in f_j is quantified by $-\mu f_l E_j^l$.
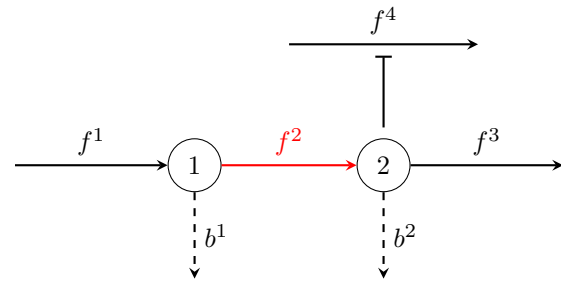

Fig. 3. The dependence of marginal opportunity on the reaction neighborhood. The figure shows a simple example of a reaction ($j = 2$, red) that is directly connected to two metabolites ($m = 1, 2$) and thereby to two other reactions $j = 1, 3$. Reaction $j = 2$ is also connected indirectly to reaction $j = 4$ by inhibiting it through metabolite $m = 2$ (indicated by the blunt arrow ⊥). The marginal opportunity of $j = 2$ is $-\mu f_l E_2^l$, where $E_2^l = \frac{\partial \tau^l}{\partial b^i}M_2^i = \frac{\partial \tau^l}{\partial b^1}M_2^1 + \frac{\partial \tau^l}{\partial b^2}M_2^2$ is determined by a marginal change in f^2 while keeping all other f^l fixed, causing (i) an inevitable change in b^1, b^2 due to the flux balance (Eq. 11); which by consequence causes (ii) an inevitable change in $\tau^1, \tau^2, \tau^3, \tau^4$, as these are functions of b^1, b^2; which finally causes (iii) an inevitable change in p^1, p^2, p^3, p^4 due to the kinetic constraints (Eq. 4) at fixed v^1, v^2, v^3, v^4 (determined by the fixed flux fractions and growth rate (Eq. 10)). The example also shows how the information about mass conservation and reaction kinetics is completely built into the definition of the growth function (Eq. 16).

The above results confirm the often postulated central role of proteins in the cellular economy (33–35). While the measure of cellular economic value may be the growth rate itself, protein concentrations constitute the general currency in which we can express the contributions of cellular subsystems. We can highlight this central economic role of proteins further by relating the marginal values $\partial_j \mu$ – changes in growth rate in response to flux changes – to changes in the allocation of proteome fractions $\phi^l := p^l / c^{\mathrm{a}}$:

$$\partial_j \mu = \frac{\mu}{b^{\mathrm{a}}} \left(M_j^{\mathrm{a}} - \mu \tau_j - \mu f_l E_j^l \right) = -\mu \sum_l \left(\frac{\partial \phi^l}{\partial f^j} \right)_\mu \quad . \tag{43}$$

The second equality follows directly from taking the derivative of $\sum_l \phi^l = \mu \rho \tau_l f^l / \rho M_l^{\mathrm{a}} f^l$ with respect to f^j at constant μ.

We can now look at the balance equations at optimal growth from an economic perspective. For the ribosome and active enzymatic reactions, a zero marginal value $\partial_j \mu = 0$ also means a zero shadow price, $\lambda_j = 0$ (Eqs. 33, 34) – so the reaction is optimal, and growth cannot be accelerated by increasing or decreasing f^j by a small amount. This insight provides an intuitive interpretation for the balance equations for the ribosome (Eq. 36) and for all active, internal enzymatic reactions (Eq. 37). An exception are only the transporters. In contrast to all other flux fractions, their shadow price (Eq. 35) depends both on their marginal value and on their *marginal biomass production*, $\mu f_l E_j^l f^j \gamma_s$ (a cost when negative and a benefit when positive).

For active enzymes with zero marginal value – and thus for all active enzymes at optimality (Eq. 37) – Eq. 31 simplifies to

$$\tau_e + f_l E_e^l = 0 \quad . \tag{44}$$

This simple relationship shows that at optimality, the marginal investment into e should perfectly balance its marginal opportunity. As the last equation involves only the neighborhood of e (defined as all reactions l such that $E_e^l \neq 0$), we can study such relationships at optimality locally, without full knowledge about the entire reaction network. We thus do not need the entire matrix $\mathbf{M}$ or complete knowledge of parameterized turnover time functions in the vector $\boldsymbol{\tau}$.

In the preceding subsection, we studied how any optimal or non-optimal growth rate μ is sensitive to marginal changes in one of the flux fractions, resulting in an economic understanding of marginal flux values in terms of their relationship with protein allocation. We next reinterpret some of these results from the perspective of control theory, and turn to a complementary problem that focuses on the sensitivity of the optimal growth rate to changes in the model parameters and external concentrations.

Growth Control and Adaptation. We are first interested in the total control that each f^j has on the (optimal or non-optimal) growth rate μ, accounting also for the density constraint limitation. In order to do that, we choose one active transport reaction s' and express its corresponding $f^{s'} \neq 0$ as a function of the other fluxes via the density constraint (Eq. 18),

$$f_{s'} = \frac{1}{\gamma_{s'}} \left(1 - \gamma_l f^l \right) \quad , \tag{45}$$

where $l \neq s'$ sums over all other reactions. Thus,

$$\frac{\partial f^{s'}}{\partial f^j} = -\frac{\gamma_j}{\gamma_{s'}} \quad , \tag{46}$$

which is non-zero only if j is also a transport reaction (so $\gamma_j \neq 0$). We now define the *Growth Control Coefficients* Γ_j as

$$\Gamma_j := \partial_j \mu - \partial_{s'} \mu \frac{\gamma_j}{\gamma_{s'}} \quad , \tag{47}$$

where the first term quantifies the growth change caused by f^j itself, and the second term quantifies the growth change caused by a change in $f^{s'}$, itself changed due to the changed f^j and the density constraint. Note that for the ribosome and enzymatic reactions, their growth control coefficient is simply their marginal value, since $\gamma_{\mathrm{r}} = \gamma_e = 0$. For models with only one transport reaction s, $\Gamma_s = 0$, since $f_s = (\gamma_s)^{-1}$ is fixed by the density constraint and cannot be changed. Conveniently, this is also captured by Eq. 47. If s' is optimal, $\lambda_{s'} = 0$, and Eq. 28 determines $\partial_{s'} \mu = -\lambda_\rho \gamma_{s'}$, so in that case

$$\Gamma_j = \partial_j \mu + \lambda_\rho \gamma_j \quad , \tag{48}$$

and Eqs. 36, 37, 38 are thus equivalent to

$$\Gamma_j f_j = 0 \quad . \tag{49}$$

We may also see the optimal condition for enzymes (Eq. 44) in terms of protein concentrations, by multiplying it element-wise with v_e (so it is also valid now for inactive enzymes),

$$p_e = p_l C_e^l \quad , \tag{50}$$

where we defined

$$C_e^l := \frac{f_e}{v^l} \left(\frac{\partial v^l}{\partial f^e} \right)_{p^l} = \frac{f_e}{v^l} \left(\frac{\partial (p^l / \tau^l)}{\partial f^e} \right)_{p^l} = -\frac{f_e}{\tau^l} \frac{\partial \tau^l}{\partial f^e} \quad , \tag{51}$$

via Eq. 4 and using partial derivatives at fixed p^l. C_e^l can been seen as (scaled) *control coefficients* (CC), analogous to (scaled) control coefficients in MCA (9, 10). This result is analogous to how enzyme concentrations and their respective CC relate at optimal fluxes constrained by a fixed total enzyme concentration (36) (see SI text "Optimal enzyme concentrations and control coefficients" for a detailed discussion). For an example of control coefficients where τ follows a simple Michaelis-Menten rate law, see SI text "Rate laws and kinetic parameters".

We now explore the sensitivity of the optimal growth rate to changes in one parameter π in the vector $\boldsymbol{\pi}$. The growth problem (Eq. 21) is constrained by the parameters $\boldsymbol{\pi}$, including the arguments necessary to determine the turnover times $\boldsymbol{\tau}$ at given $\mathbf{f}$. This means that any marginal change in one of the parameters $\boldsymbol{\pi}$ would lead to changes in the solution $\mathbf{f}^*$ of the

optimization (Eq. 21). In this sense, the parameters $\boldsymbol{\pi}$ can be understood as control variables, while the corresponding optimal state $\mathbf{f}^*$, and its functions $\mu^* = \mu(\mathbf{f}^*), \mathbf{v}^* = \mathbf{v}(\mathbf{f}^*), \mathbf{p}^* = \mathbf{p}(\mathbf{f}^*)$, and $\mathbf{c}^* = \mathbf{c}(\mathbf{f}^*)$ are the response variables. Figure 4 summarizes these relationships.

$$\boldsymbol{\pi} \xrightarrow{\text{optimization (21)}} \mathbf{f}^* \xrightarrow{Eqs.\ 4,9,10,16} \mathbf{p}^*, \mathbf{c}^*, \mathbf{v}^*, \mu^*$$

Fig. 4. The parameters $\boldsymbol{\pi}$ and their control on the optimal cellular state $\mathbf{f}^*$.

Because growth rate is closely related to fitness, we are also particularly interested in how marginal changes in one of the previously fixed parameters $\boldsymbol{\pi}$ affect the *optimal* growth rate μ^* (37). We can estimate this effect directly via the envelope theorem (4, 38), by effectively considering the *optimal* state $\mathbf{f}^*$ as fixed and treating the parameters $\boldsymbol{\pi}$ as the new independent variables, making it unnecessary to calculate the new optimal state after the parameter change. To do that, we first simplify the problem by assuming that these marginal changes have no effect on which reactions are active, so we simplify the Lagrangian (Eq. 25) by ignoring the inequality constraints; note that in this case only the objective function μ can be influenced by parameter changes, since the density constraint only depends on $\mathbf{M}$, whose entries cannot be changed continuously. Second, we can think about the optimal growth rate μ^* as a function of the parameters $\mu^*(\boldsymbol{\pi}) := \mathcal{L}(\mathbf{f}^*(\boldsymbol{\pi}), \boldsymbol{\lambda}^*(\boldsymbol{\pi}), \boldsymbol{\pi})$, so the the total change $\mathrm{d}\mu^*/\mathrm{d}\pi$ induced by a marginal change in a parameter π can be calculated via the chain rule

$$\frac{\mathrm{d}\mu^*}{\mathrm{d}\pi} = \frac{\partial \mathcal{L}}{\partial f_j} \frac{\mathrm{d}f_j^*}{\mathrm{d}\pi} + \frac{\partial \mathcal{L}}{\partial \lambda_\rho} \frac{\mathrm{d}\lambda_\rho^*}{\mathrm{d}\pi} + \frac{\partial \mathcal{L}}{\partial \pi} = \frac{\partial \mathcal{L}}{\partial \pi} \quad , \qquad \textbf{(52)}$$

where the last equality comes from $\partial \mathcal{L}/\partial f^j = \partial \mathcal{L}/\partial \lambda_\rho = 0$ according to the stationarity (Eq. 26) and primal feasibility (Eq. 18) at an optimal state.

We now define *growth adaptation coefficients* A as the relative change in the *optimal* growth rate μ^* in response to a small, relative change in one control variable π

$$A_\pi := \frac{\pi}{\mu^*} \frac{\mathrm{d}\mu^*}{\mathrm{d}\pi} = \frac{\pi}{\mu(\mathbf{f})} \frac{\partial \mathcal{L}}{\partial \pi} \quad , \qquad \textbf{(53)}$$

where here and in the rest of this section $\mathbf{f}$ is to be understood as the *optimal* state before the change in the parameter π. Note that in the following discussion, the parameters π of interest only influence $\mathcal{L}$ via the objective function μ, so the partial derivatives $\partial \mathcal{L}/\partial \pi$ are simply evaluated as $\partial \mu/\partial \pi$ at fixed $\mathbf{f}$.

For direct changes in the turnover times τ^j (e.g., through changing the corresponding $1/k_{\text{cat}}^j$), the growth adaptation coefficient is calculated by evaluating the growth function μ and its partial derivative at fixed $\mathbf{f}$,

$$A_{\tau^j} := \frac{\tau_j}{\mu^*} \frac{\mathrm{d}\mu^*}{\mathrm{d}\tau^j} = \frac{\tau_j}{\mu} \frac{\partial \mu}{\partial \tau^j} = -\frac{\mu \rho f_j \tau_j}{c_a} = -\phi_j \quad , \qquad \textbf{(54)}$$

where we effectively treated τ^j as a variable in the growth Eq. 16, and $\phi_j = p_j/c_a$ is the optimal proteome fraction allocated to reaction j before the change in τ^j. This result is

consistent with the observation that drugs targeting the most highly expressed catalysts, such as the ribosome, have the strongest effects on cellular growth rates (5, 39).

For changes in some external parameter such as a concentration x, the growth adaptation coefficient is again calculated by evaluating the growth function μ and its partial derivative at fixed $\mathbf{f}$, and using the chain rule of differentiation we obtain

$$A_x := \frac{x}{\mu^*} \frac{\mathrm{d}\mu^*}{\mathrm{d}x} = \frac{x}{\mu} \frac{\partial \mu}{\partial \tau^s} \frac{\partial \tau^s}{\partial x} = -\phi_s \frac{x}{\tau^s} \frac{\partial \tau^s}{\partial x} \quad , \qquad \textbf{(55)}$$

where we have a summation over s (only transporters s have kinetic rate laws depending on external concentrations). According to Eq. 55, the growth adaptation coefficient of an external concentration x is simply the sum over the "scaled elasticities" $\dfrac{x}{\tau^s} \dfrac{\partial \tau^s}{\partial x}$ of the transporters of x, weighted by the optimal proteome fractions ϕ_s allocated to each s before the change in x. This result gives an explicit quantitative estimation on which external concentrations should be changed in order to cause the most change in the optimal growth rate. This equation may hence provide a useful tool for improving the growth media environment for industrial cell cultures, and for quantifying the effect of drugs aimed at decreasing the growth of pathogens and cancer cells. If the turnover times $\boldsymbol{\tau}$ depend explicitly on other external parameters, such as pH and temperature, growth adaptation coefficients can be calculated and interpreted exactly as in Eq. 55.

The growth adaptation coefficient with respect to the mass density ρ, assuming it affects turnover times $\boldsymbol{\tau}$ only through reactant concentrations $\mathbf{c}$, reads

$$A_\rho := \frac{\rho}{\mu^*} \frac{\mathrm{d}\mu^*}{\mathrm{d}\rho} = \frac{\rho}{\mu} \frac{\partial \mu}{\partial \tau^l} \frac{\partial \tau^l}{\partial c^i} \frac{\partial c^i}{\partial \rho} = -\frac{\rho}{\mu} \left(\frac{\mu^2}{b^a} f_l \frac{\partial \tau^l}{\partial c^i} M_j^i f^j \right)$$
$$= -\frac{\mu}{b^a} f_l E_j^l f^j \quad , \qquad \textbf{(56)}$$

where $\partial c^i/\partial \rho = b^i = M_j^i f^j$ according to Eqs. 9,11, and the last equality comes from the definition of the indirect elasticity (Eq. 24). From this expression and λ_ρ in Eq. 32, we see that $-\lambda_\rho = \mu A_\rho$; at optimality, the negative Lagrange multiplier for the density constraint, $-\lambda_\rho$, quantifies the absolute increase in growth rate caused by a marginal increase in ρ, given by μ itself times the proportional change, A_ρ. Thus, the extra term in the shadow price of transporters (compare λ_s in Eq. 35 to λ_r and λ_e) quantifies the growth rate benefit gained by allowing the violation of the density constraint (Eq. 18) caused by a small increase in f_s.

Just as the economy of growth is deeply connected to protein allocation, so is growth control. For A_j and A_x, this connection is clear from Eqs. 54, 55, respectively. For A_ρ, we first note that it relates to optimal marginal values via Eq. 30,

$$\mu A_\rho = -\lambda_\rho = (\partial_j \mu) f^j = (\partial_s \mu) f^s \quad . \qquad \textbf{(57)}$$

At optimality, the summands on the RHS are zero for the ribosome and for enzymatic reactions ($\partial_j \mu f_j = 0$ for $j = \mathrm{r}, e$), and the summation over j can thus be restricted to only transporters s. Thus, at optimality, the absolute change in optimal growth rate caused by increasing ρ, μA_ρ, is equal to the

summed marginal effects of transport fluxes on the growth rate, $\partial_s \mu$, weighted by the flux fractions f^s themselves. To see the full connection between A_ρ and protein allocation, we insert Eq. 43 into Eq. 57 to obtain

$$A_\rho = -\sum_l \left(\frac{\partial \phi^l}{\partial f^s} \right)_\mu f^s \quad , \qquad (58)$$

where we sum over s. This equation shows that the proportional effect on the optimal growth rate that is exerted by a marginal increase in ρ, A_ρ, equals the combined marginal effects of transport fluxes f^s on proteome allocation fractions, weighted by the transport fluxes themselves.

Discussion

Modeling frameworks that are essentially linear, such as FBA and RBA, are typically analyzed numerically, as the efficiency of linear programming facilitates fast solutions even for genome-scale models (7, 8, 11). In contrast, the construction and solution of genome-scale non-linear models faces two major obstacles, both intimately linked to the kinetic rate laws. First, experimental estimates for the required kinetic parameters – k_{cat} and K_{m} values in the simplest case of generalized Michaelis-Menten kinetics – are lacking for most reactions (40). This problem can be alleviated by using parameter estimates from artificial intelligence approaches (41–43). Second, the non-linearity of enzymatic rate laws makes numerical optimizations much more difficult than for linear systems, explaining why existing studies have been limited to models with only a handful of reactions (6, 15–19). Numerical optimization is particularly problematic for models with redundant pathways, where the optimization problem is non-convex (20).

The succinct mathematical formulation for modeling balanced cellular growth developed in this paper helps to address both problems. On the one hand, the reduction of the problem description to a minimal number of independent variables – the flux fractions – reduces the dimensionality of the search space, and can thus accelerate numerical approaches to find optimal states. On the other hand, this formulation allowed us to identify necessary conditions for states of maximal growth rate. These conditions are local for each reaction, i.e., they do not require complete knowledge of the cellular reaction network and its kinetics. In particular, these local balance equations provide a necessary condition for the activity of each individual reaction at optimality: for the corresponding flux to be nonzero, the terms in parentheses in Eqs. 36-38 have to be zero. Our analytical approach thus provides a tool for a deeper understanding of the general principles shaping optimal cellular resource allocation, even when specific optimal states are not explicitly known.

The concise formulation also helped in the interpretation of the optimality conditions from the perspectives of economy and control theory. The marginal change in growth rate induced by each flux change is seen as the flux's marginal economic value, while the growth adaptation coefficient of each model parameter or external concentration is the change in the optimal growth rate induced by a marginal change in this parameter. The close correspondence between the mathematical expressions obtained in both perspectives helps to clarify the mathematical and conceptual links between these usually separate fields of study, including the extension of previous results of metabolic control analysis (MCA), developed for ad-hoc objectives in static sub-networks, to the holistic problem of cellular growth in GBA models. In MCA, one typically treats enzyme concentrations as control variables and studies how small changes to them affect reactant concentrations and fluxes. Here, all these variables are not only connected, but are uniquely determined by the flux fraction vector $\mathbf{f}$. Moreover, the growth rate μ itself is explicitly connected to $\mathbf{f}$ through the growth function (Eq. 16). Through these connections, we can quantify the sensitivity of the cellular growth rate, and hence approximately of organismal fitness, to changes in the control variables $\boldsymbol{\pi}$, something not possible in the usual MCA framework (9, 10). The growth adaptation coefficients provide explicit expressions for the effects on growth rate caused by small changes in control variables at optimality. Due to the close relationship between growth rate and fitness, these estimates could be used to interpret and predict evolutionary changes in these variables.

A closely related nonlinear cellular modeling approach accounts for the different amino acid compositions of individual proteins by including "personalized" ribosome reactions for each protein (44–46). In contrast to GBA, this type of model cannot be simplified using flux fractions $\mathbf{f}$, as it requires a mathematical formulation that includes explicit variables for metabolite concentrations. Experimental data for *E. coli* (47) indicates that the 20 amino acid content into its total proteome changes very little over 22 highly distinct growth environments (mean coefficient of variation = 2.46%, maximal CV = 7.55%, see Table S1), suggesting that – at least globally – different protein compositions are likely not a major factor driving significant changes in the optimal cellular state. Thus, a unique ribosome reaction with fixed column M_{r} is a realistic assumption over all these growth conditions. Further study is necessary to identify whether the different compositions of individual proteins may cause significant changes in their allocation across environments.

All analytical results in this study were derived exclusively from the growth constraints assumed in GBA models: mass conservation in balanced growth, reaction kinetics, cellular density, and non-negative concentrations. For the analysis of optimal growth, we encoded all corresponding information into a single Lagrangian function, parameterized in terms of the constraints. We formulated the problem with the flux fractions $\mathbf{f}$ as the only free variables, and used KKT conditions to obtain the necessary conditions for optimal growth states. Through these conditions, the marginal protein allocation emerges as the natural underlying currency in the cell economy; this relationship has frequently been asserted (33–35), but is derived here entirely from first principles.

The KKT framework provides a straight-forward way to incorporate new constraints, analogous to how physical theories using the Lagrangian formalism account for additional

forces by adding corresponding functions and Lagrange multipliers into the Lagrangian. A re-derivation of the KKT conditions will then result in an extended set of balance equations. Among the potential extra physiological constraints, one might consider also phenomenological constraints such as the recently reported relationship between the cellular surface/volume ratio and the growth rate (48).

One fundamental physiological limitation that could be included in this way but is not considered explicitly here is the diffusion limit of molecules within cellular compartments. This limit links density and kinetic constraints. A higher dry mass density increases the "crowding effect" within cells (25), which entails a lower diffusion rate and by consequence a longer time for reactants to find their catalysts; this effect can be modeled directly by including a corresponding dependence in the Michaelis constants K_{m}. A study on the crowding effects of all cellular concentrations – including those of small molecules – found that the observed *E. coli* dry mass density is in the range expected if evolution had optimized the cellular density for maximal growth rate (30). In this sense, a fixed density constraint on all molecules, as considered here, may be seen as a simplifying approximation, justified by the observed constancy of cellular buoyant and dry mass densities across different growth conditions (24, 48), with the exception only of large changes in environmental osmolarities (25).

The Lagrangian formalism described here also allows a direct generalization of the theory to other objective functions, i.e., other measures of fitness at balanced growth. This can be done by incorporating a new objective function $F(\mathbf{f})$ and adding a new constraint for the growth rate via $\lambda_\mu(\mu - \mu_0)$, where μ is determined by the growth function, λ_μ is the corresponding KKT multiplier, and μ_0 is the constrained growth rate given now as an input.

An important step toward a more general theory of cellular growth would be to extend the GM theory to changing environments, and to derive similar analytical conditions for time-dependent optimal cellular states $\mathbf{f}(t)$. In this situation, fitness is determined by the total growth in a given period of time, so the objective function becomes the integral of the specific growth rate (17, 49), under the same constraints as discussed here. This dynamical extension of the theory might also consider the biologically important scenario of cyclical environments, such as feast-famine cycles of the gut microbiome (50) or day-night cycles of photosynthetic microbes (19).

In sum, the concise mathematical formulation of the growth optimization problem developed here provides a powerful toolbox for the analysis and solution of mechanistic descriptions of optimal cellular physiology and growth. It thereby opens a path toward a fundamental understanding of organizing principles of biological cells. While biological systems will never be fully optimal, optimal states provide an extremely useful null model for the action of natural selection.

Methods

Let us consider a Lagrangian including inequality constraints on metabolite concentrations,

$$\mathcal{L}(\mathbf{f}, \boldsymbol{\lambda}) = \mu(\mathbf{f}) + \lambda_\rho \left(\gamma_s f^s - 1 \right) + \lambda_l f^l \tau^l(\mathbf{f}) + \lambda_m M_l^m f^l \quad. \tag{59}$$

The necessary KKT conditions are the same as in the main text, plus extra conditions for the constraints on metabolite concentrations, including

$$M_j^m f^j \geq 0 \quad \text{(primal feasibility)} \tag{60}$$

$$\lambda^m M_j^m f^j = 0 \quad \text{(complementary slackness)} \tag{61}$$

The last equation corresponds to $\lambda^m b^m = 0$, but since cells in optimal states will only express those proteins that are actually needed to catalyze reactions, we can be even more restrictive and require a stronger version of the last equation:

$$\lambda^m M_j^m f_j = 0 \quad. \tag{62}$$

This equation encodes the following: if $\lambda^m \neq 0$, the metabolite m is inactive ($c^m = 0$), and then all of the reactions j connected to it (j such that $M_j^m \neq 0$) must also be inactive ($f_j = 0$). We also note that because the turnover times $\boldsymbol{\tau}$ differ from zero, the complementary slackness of reaction j (Eq. 27) is equivalent to

$$\lambda_j f_j = 0 \quad. \tag{63}$$

Now we solve the necessary equality conditions, first considering the stationarity

$$\partial_j \mathcal{L} = \partial_j \mu + \lambda_\rho \gamma_j + \lambda_l f^l E_j^j + \lambda_j \tau_j + \lambda_m M_j^m = 0 \quad. \tag{64}$$

By considering the stronger version of the complementary slackness for reactions (Eq. 63), we cancel the sum on λ_l, resulting in

$$\partial_j \mu + \lambda_\rho \gamma_j + \lambda_j \tau_j + \lambda_m M_j^m = 0 \quad. \tag{65}$$

Now we multiply (element-wise) by f_j

$$(\partial_j \mu) f_j + \lambda_\rho \gamma_j f_j + \lambda_j \tau_j f_j + \lambda_m M_j^m f_j = 0 \tag{66}$$

and consider the stronger version of the complementary slackness (Eq. 62) on metabolites, and the complementary slackness on reactions (Eq. 27), resulting in Eq. 29. Summing over all j and using the density constraint (Eq. 18) results in Eq. 30. Now the complete expression for λ_ρ in terms of $\mathbf{f}$ is given by substituting the expression for the growth rate derivatives (Eq. 31),

$$\lambda_\rho = -\frac{\mu}{b^{\mathrm{a}}} \left(M_j^{\mathrm{a}} f^j - \mu \tau_j f^j - \mu f_l E_j^l f^j \right) \quad. \tag{67}$$

The first term in parentheses equals b^{a} according to Eq. 15, and the third equals $-b^{\mathrm{a}}$ according to Eq. 14, so that both terms cancel each other. This results exactly in Eq. 32.

Finally, combining Eqs. 29, 32, we obtain a general form for the balance equations,

$$\left(\partial_j \mu - \frac{\mu^2}{b^{\mathrm{a}}} f_l E_h^l f^h \gamma_j \right) f_j = 0 \quad, \tag{68}$$

where both indices l and h are used to indicate summation over all reactions. Using Eq. 31 for the derivative, we have

$$\left(M_j^{\mathrm{a}} - \mu\tau_j - \mu\, f_l E_j^l + \mu\, f_l E_h^l f^h \gamma_j \right) f_j = 0 \quad . \qquad \textbf{(69)}$$

We use Eqs. 9,10,11 to express this in terms of $\mathbf{v}, \mathbf{c}$,

$$\left(M_j^{\mathrm{a}} - \mu\tau_j - v_l \frac{\partial \tau^l}{\partial c^i} M_j^i + v_l \frac{\partial \tau^l}{\partial c^i} \frac{c^i}{\rho} \gamma_j \right) v_j = 0 \quad , \qquad \textbf{(70)}$$

Using $v_l = p_l/\tau_l$ from Eq. 4, we obtain

$$\left(M_j^{\mathrm{a}} - \mu\tau_j - \frac{p_l}{\tau_l} \frac{\partial \tau^l}{\partial c^i} M_j^i + \frac{p_l}{\tau_l} \frac{\partial \tau^l}{\partial c^i} \frac{c^i}{\rho} \gamma_j \right) v_j = 0 \quad . \qquad \textbf{(71)}$$

By multiplying out v_j, we get Eq. 39.

Bibliography

1. Dennis P Bremer H. Modulation of chemical composition and other parameters of the cell at different exponential growth rates. *EcoSal Plus*, 2008.

2. R. A. Fisher and J. H. Bennett. *The genetical theory of natural selection: a complete variorum edition*. Oxford University Press, 1999. ISBN 978-0-19-850440-5.

3. Hugo Dourado, Matteo Mori, Terence Hwa, and Martin J. Lercher. On the optimality of the enzyme–substrate relationship in bacteria. *PLOS Biology*, 19(10):1–18, 10 2021. doi: 10.1371/journal.pbio.3001416.

4. Hugo Dourado and Martin J. Lercher. An analytical theory of balanced cellular growth. *Nature Communications*, 11(1):1226, Mar 2020. ISSN 2041-1723. doi: 10.1038/s41467-020-14751-w.

5. Xiao-Pan Hu, Hugo Dourado, Peter Schubert, and Martin J. Lercher. The protein translation machinery is expressed for maximal efficiency in escherichia coli. *Nature Communications*, 11(1):5260, Oct 2020. ISSN 2041-1723. doi: 10.1038/s41467-020-18948-x.

6. Douwe Molenaar, Rogier van Berlo, Dick de Ridder, and Bas Teusink. Shifts in growth strategies reflect tradeoffs in cellular economics. *Molecular Systems Biology*, 5(1):323, 2009. doi: 10.1038/msb.2009.82.

7. Anne Goelzer, Vincent Fromion, and Gérard Scorletti. Cell design in bacteria as a convex optimization problem. *Automatica*, 47(6):1210–1218, 2011. ISSN 0005-1098. doi: 10.1016/j.automatica.2011.02.038.

8. Edward J. O'Brien, Joshua A. Lerman, Roger L. Chang, Daniel R. Hyduke, and Bernhard Palsson. Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction. *Molecular Systems Biology*, 2013. ISSN 17444292. doi: 10.1038/msb.2013.52.

9. Reinhart Heinrich and Tom A. Rapoport. A linear steady-state treatment of enzymatic chains. *European Journal of Biochemistry*, 42(1):89–95, 1974. doi: https://doi.org/10.1111/j.1432-1033.1974.tb03318.x.

10. H. Kacser and J. A. Burns. The control of flux. *Symp Soc Exp Biol*, 27:65–104, 1973.

11. Nathan E. Lewis, Harish Nagarajan, and Bernhard O. Palsson. Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods. *Nature Reviews Microbiology*, 10(4):291–305, 2012. ISSN 17401534. doi: 10.1038/nrmicro2737.

12. M. R. Watson. Metabolic maps for the Apple II. *Biochemical Society Transactions*, 12(6):1093–1094, 1984. ISSN 0300-5127. doi: 10.1042/bst0121093.

13. Anne Goelzer, Jan Muntel, Victor Chubukov, Matthieu Jules, Eric Prestel, Rolf Nölker, Mahendra Mariadassou, Stéphane Aymerich, Michael Hecker, Philippe Noirot, Dörte Becher, and Vincent Fromion. Quantitative prediction of genome-wide resource allocation in bacteria. *Metabolic Engineering*, 32:232–243, 2015. ISSN 10967176. doi: 10.1016/j.ymben.2015.10.003.

14. Matteo Mori, Terence Hwa, Olivier C. Martin, Andrea De Martino, and Enzo Marinari. Constrained allocation flux balance analysis. *PLOS Computational Biology*, 12(6):1–24, 2016. doi: 10.1371/journal.pcbi.1004913.

15. Andrea Y. Weiße, Diego A. Oyarzún, Vincent Danos, and Peter S. Swain. Mechanistic links between cellular trade-offs, gene expression, and growth. *Proceedings of the National Academy of Sciences*, 112(9):E1038–E1047, 2015. ISSN 0027-8424. doi: 10.1073/pnas.1416533112.

16. Arijit Maitra and Ken A. Dill. Bacterial growth laws reflect the evolutionary importance of energy efficiency. *Proceedings of the National Academy of Sciences*, 112(2):406–411, 2015. ISSN 0027-8424. doi: 10.1073/pnas.1421138111.

17. Nils Giordano, Francis Mairet, Jean-Luc Gouzé, Johannes Geiselmann, and Hidde de Jong. Dynamical Allocation of Cellular Resources as an Optimal Control Problem: Novel Insights into Microbial Growth Strategies. *PLOS Computational Biology*, 12(3):e1004802, 2016. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1004802.

18. Benjamin D Towbin, Yael Korem, Anat Bren, Shany Doron, Rotem Sorek, and Uri Alon. Optimality and sub-optimality in a bacterial growth law. *Nature Communications*, 8:14123, 2017. ISSN 20411723. doi: 10.1038/ncomms14123.

19. Marjan Faizi, Tomáš Zavřel, Cristina Loureiro, Jan Červený, and Ralf Steuer. A model of optimal protein allocation during phototrophic growth. *BioSystems*, 166:26–36, apr 2018. ISSN 18728324. doi: 10.1016/j.biosystems.2018.02.004.

20. Meike T. Wortel, Elad Noor, Michael Ferris, Frank J. Bruggeman, and Wolfram Liebermeister. Metabolic enzyme cost explains variable trade-offs between microbial growth rate and yield. *PLOS Computational Biology*, 14(2):1–21, 02 2018. doi: 10.1371/journal.pcbi.1006010.

21. Stefan Müller, Georg Regensburger, and Ralf Steuer. Enzyme allocation problems in kinetic metabolic networks: Optimal solutions are elementary flux modes. *Journal of Theoretical Biology*, 347:182 – 190, 2014. ISSN 0022-5193. doi: https://doi.org/10.1016/j.jtbi.2013.11.015.

22. Meike T. Wortel, Han Peters, Josephus Hulshof , Bas Teusink, and Frank J. Bruggeman. Metabolic states with maximal specific rate carry flux through an elementary flux mode. *The FEBS Journal*, 281(6):1547–1555, 2014. doi: 10.1111/febs.12722.

23. W. W. Baldwin, Richard Myer, Nicole Powell, Erika Anderson, and Arthur L. Koch. Buoyant density of escherichia coli is determined solely by the osmolarity of the culture medium. *Archives of Microbiology*, 164(2):155–157, 1995. ISSN 1432-072X. doi: 10.1007/s002030050248.

24. H E Kubitschek, W W Baldwin, and R Graetzer. Buoyant density constancy during the cell cycle of escherichia coli. *Journal of bacteriology*, 155:3, 9 1983.

25. Scott Cayley, Barbara A. Lewis, Harry J. Guttman, and M.Thomas Record. ...characterization of the cytoplasm of escherichia coli k-12 as a function of external osmolarity: Implications for protein-dna interactions in vivo. *Journal of Molecular Biology*, 222(2):281 – 300, 1991. ISSN 0022-2836. doi: https://doi.org/10.1016/0022-2836(91)90212-O.

26. Jerry B. Marion and Stephen T. Thornton. *Classical dynamics of particles and systems; 3rd edition*. Harcourt Brace Jovanovich, San Diego, 1988. ISBN 015507640X.

27. David G Luenberger and Yinyu Ye. *Linear and nonlinear programming*. International Series in Operations Research Management Science. Springer, New York, NY, 3 edition, July 2008.

28. Wolfram Liebermeister. Optimal metabolic states in cells. *bioRxiv*, 2018. doi: 10.1101/483867.

29. Wolfram Liebermeister. The value structure of metabolic states. *bioRxiv*, 2018. doi: 10.1101/483891.

30. Tin Yau Pang and Martin J. Lercher. Optimal density of biological cells. *bioRxiv*, 2020. doi: 10.1101/2020.11.18.388744.

31. William Karush. Minima of functions of several variables with inequalities as side constraints. *M. Sc. Dissertation. Dept. of Mathematics, Univ. of Chicago*, 1939.

32. Harold W Kuhn and Albert W Tucker. Nonlinear programming proceedings of the second berkeley symposium on mathematical statistics and probability. *Neyman*, pages 481–492, 1951.

33. Erez Dekel and Uri Alon. Optimality and evolutionary tuning of the expression level of a protein. *Nature*, 436(7050):588–592, 2005. ISSN 1476-4687. doi: 10.1038/nature03842.

34. Elad Noor, Avi Flamholz, Arren Bar-Even, Dan Davidi, Ron Milo, and Wolfram Liebermeister. The protein cost of metabolic fluxes: Prediction from enzymatic rate laws and cost minimization. *PLOS Computational Biology*, 12(11):1–29, 2016. doi: 10.1371/journal.pcbi.1005167.

35. Moshe Kafri, Eyal Metzl-Raz, Ghil Jona, and Naama Barkai. The Cost of Protein Production. *Cell Reports*, 14(1):22–31, 2016. ISSN 22111247. doi: 10.1016/j.celrep.2015.12.015.

36. Edda Klipp and Reinhart Heinrich. Competition for enzymes in metabolic pathways:: Implications for optimal distributions of enzyme concentrations and for the distribution of flux control. *Biosystems*, 54(1):1 – 14, 1999. ISSN 0303-2647. doi: https://doi.org/10.1016/S0303-2647(99)00059-3.

37. St Elmo Wilken, Mathieu Besançon, Miroslav Kratochvíl, Chilperic Armel Foko Kuate, Christophe Trefois, Wei Gu, and Oliver Ebenhöh. Interrogating the effect of enzyme kinetics on metabolism using differentiable constraint-based models. *Metabolic engineering*, page S1096—7176(22)00117—3, September 2022. ISSN 1096-7176. doi: 10.1016/j.ymben.2022.09.002.

38. S. Afriat. Theory of maxima and the method of lagrange. *SIAM Journal on Applied Mathematics*, 20(3):343–357, 1971. doi: 10.1137/0120037.

39. Matthew Scott, Carl W. Gunderson, Eduard M. Mateescu, Zhongge Zhang, and Terence Hwa. Interdependence of cell growth and gene expression: Origins and consequences. *Science*, 330(6007):1099–1102, 2010. ISSN 00368075. doi: 10.1126/science.1192588.

40. Avlant Nilsson, Jens Nielsen, and Bernhard O. Palsson. Metabolic models of protein allocation call for the kinetome. *Cell Systems*, 5(6):538–541, 2017. ISSN 2405-4712. doi: 10.1016/j.cels.2017.11.013.

41. David Heckmann, Colton J. Lloyd, Nathan Mih, Yuanchi Ha, Daniel C. Zielinski, Zachary B. Haiman, Abdelmoneim Amer Desouki, Martin J. Lercher, and Bernhard O. Palsson. Machine learning applied to enzyme turnover numbers reveals protein structural correlates and improves metabolic models. *Nature Communications*, 9(1):5252, 2018. ISSN 2041-1723. doi: 10.1038/s41467-018-07652-6.

42. Alexander Kroll, Martin K. M. Engqvist, David Heckmann, and Martin J. Lercher. Deep learning allows genome-scale prediction of michaelis constants from structural features. *PLOS Biology*, 19(10):1–21, 10 2021. doi: 10.1371/journal.pbio.3001402.

43. Feiran Li, Le Yuan, Hongzhong Lu, Gang Li, Yu Chen, Martin K. M. Engqvist, Eduard J. Kerkhoven, and Jens Nielsen. Deep learning-based kcat prediction enables improved enzyme-constrained model reconstruction. *Nature Catalysis*, 5(8):662–672, August 2022. ISSN 2520-1158. doi: 10.1038/s41929-022-00798-z. Number: 8 Publisher: Nature Publishing Group.

44. Daan H. de Groot, Josephus Hulshof, Bas Teusink, Frank J. Bruggeman, and Robert Planqué. Elementary growth modes provide a molecular description of cellular self-fabrication. *PLOS Computational Biology*, 16(1):1–33, 01 2020. doi: 10.1371/journal.pcbi.1007559.

45. Stefan Müller. Elementary growth modes/vectors and minimal autocatalytic sets for kinetic/constraint-based models of cellular growth. *bioRxiv*, 2021. doi: 10.1101/2021.02.24.432769.

46. Stefan Müller, Diana Széliová, and Jürgen Zanghellini. Elementary vectors and autocatalytic

sets for resource allocation in next-generation models of cellular growth. *PLOS Computational Biology*, 18(2):1–27, 02 2022. doi: 10.1371/journal.pcbi.1009843.

47. Alexander Schmidt, Karl Kochanowski, Silke Vedelaar, Erik Ahrné, Benjamin Volkmer, Luciano Callipo, Kèvin Knoops, Manuel Bauer, Ruedi Aebersold, and Matthias Heinemann. The quantitative and condition-dependent escherichia coli proteome. *Nature Biotechnology*, 34:104 EP –, 2015.

48. Enno R. Oldewurtel, Yuki Kitahara, and Sven van Teeffelen. Robust surface-to-mass coupling and turgor-dependent cell width determine bacterial dry-mass density. *Proceedings of the National Academy of Sciences*, 118(32):e2021416118, 2021. doi: 10.1073/pnas.2021416118.

49. Niclas Nordholt, Johan H. van Heerden, and Frank J. Bruggeman. Biphasic cell-size and growth-rate homeostasis by single bacillus subtilis cells. *Current Biology*, 30(12):2238–2247.e5, 2020. ISSN 0960-9822. doi: https://doi.org/10.1016/j.cub.2020.04.030.

50. Matteo Mori, Severin Schink, David W. Erickson, Ulrich Gerland, and Terence Hwa. Quantifying the benefit of a proteome reserve in fluctuating environments. *Nature Communications*, 8(1):1225, 2017. ISSN 2041-1723. doi: 10.1038/s41467-017-01242-8.

51. Wolfram Liebermeister and Edda Klipp. Bringing metabolic networks to life: convenience rate law and thermodynamic constraints. *Theoretical Biology and Medical Modelling*, 3(1): 41, 2006. ISSN 1742-4682. doi: 10.1186/1742-4682-3-41.

52. S B Zimmerman and S O Trach. Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of escherichia coli. *J Mol Biol*, 222(3):599–620, December 1991.

53. Nuala A. O'Leary, Mathew W. Wright, J. Rodney Brister, Stacy Ciufo, Diana Haddad, Rich McVeigh, Bhanu Rajput, Barbara Robbertse, Brian Smith-White, Danso Ako-Adjei, Alexander Astashyn, Azat Badretdin, Yiming Bao, Olga Blinkova, Vyacheslav Brover, Vyacheslav Chetvernin, Jinna Choi, Eric Cox, Olga Ermolaeva, Catherine M. Farrell, Tamara Goldfarb, Tripti Gupta, Daniel Haft, Eneida Hatcher, Wratko Hlavina, Vinita S. Joardar, Vamsi K. Kodali, Wenjun Li, Donna Maglott, Patrick Masterson, Kelly M. McGarvey, Michael R. Murphy, Kathleen O'Neill, Shashikant Pujar, Sanjida H. Rangwala, Daniel Rausch, Lillian D. Riddick, Conrad Schoch, Andrei Shkeda, Susan S. Storz, Hanzhen Sun, Francoise Thibaud-Nissen, Igor Tolstoy, Raymond E. Tully, Anjana R. Vatsan, Craig Wallin, David Webb, Wendy Wu, Melissa J. Landrum, Avi Kimchi, Tatiana Tatusova, Michael DiCuccio, Paul Kitts, Terence D. Murphy, and Kim D. Pruitt. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Research*, 44(D1):D733–D745, 11 2015. ISSN 0305-1048. doi: 10.1093/nar/gkv1189.

Supplementary texts

Rate laws and kinetic parameters. For simplicity, it may be convenient to assume that each component τ^j in the vector of turnover times $\boldsymbol{\tau}$ has a general functional form that depends only on a set $\mathbb{K}$ of kinetic parameters relating reactions j with metabolites m and external reactants n. The simplest general rate law would be the irreversible Michaelis-Menten kinetics, which for a reaction l determines

$$\tau^l(\mathbf{c},\mathbf{x}) = \frac{1}{k^l(\mathbf{c},\mathbf{x})} = \frac{1}{k_{\text{cat}}^l} \prod_m \left(1 + \frac{K_m^l}{c_m}\right) \prod_n \left(1 + \frac{K_n^l}{x_n}\right) \quad, \tag{72}$$

where the kinetic parameters are the corresponding turnover number k_{cat}^l, and Michaelis constants K_m^l for metabolites m, and K_n^l for external reactants n. We may consider that all metabolites m and external reactants n that are not substrates in reaction l have corresponding Michaelis constants equal to zero, so the above equation doesn't depend on their respective concentrations. Note only transport reactions s depend on external concentrations $\mathbf{x}$, so $K_n^e = 0$ for all e, n, and $K_n^r = 0$ for all n.

In that case of all reactions following the irreversible Michaelis-Menten kinetics (72), the (direct) elasticity ε_m^l with respect to a metabolite m is

$$\varepsilon_m^l = \frac{\partial \tau^l}{\partial b^m} = \rho \frac{\partial \tau^l}{\partial c^m} = -\prod_{m' \neq m} \left(1 + \frac{K_{m'}^l}{c_{m'}}\right) \prod_n \left(1 + \frac{K_n^l}{x_n}\right) \frac{\rho K_m^l}{k_{\text{cat}}^l (c_m)^2} \quad, \tag{73}$$

where m' are metabolites different than m. The corresponding indirect elasticity E_j^l with respect to a reaction j is

$$E_j^l = \frac{\partial \tau^l}{\partial f^j} = \varepsilon_m^l M_j^m = -\prod_{m' \neq m} \left(1 + \frac{K_{m'}^l}{c_{m'}}\right) \prod_n \left(1 + \frac{K_n^l}{x_n}\right) \frac{\rho K_m^l}{k_{\text{cat}}^l (c_m)^2} M_j^m \quad. \tag{74}$$

where we note $\varepsilon_i^l M_j^i = \varepsilon_m^l M_j^m$ since here $\varepsilon_a^l = 0$ (the Michaelis-Menten kinetics (72) doesn't depend on the total protein concentration c^a). When $j = e$ is an enzymatic reaction, the scaling of E_e^l by the respective $-f^e$ and τ^l provides the corresponding control coefficient (Eq. 50)

$$C_e^l = -\frac{f_e}{\tau^l} \frac{\partial \tau^l}{\partial f^e} = -f_e k_{\text{cat}}^l \left(1 + \frac{K_m^l}{c_m}\right)^{-1} \frac{\rho K_m^l}{k_{\text{cat}}^l (c_m)^2} M_e^m = -f_e \frac{\rho K_m^l}{c_m(c_m + K_m^l)} M_e^m \quad. \tag{75}$$

Note that in this case the control coefficients don't depend on the turnover numbers k_{cat}, only on Michaelis constants of metabolites K_m.

A more realistic example would be some generalized kinetics such as the "convenience kinetics" proposed in Ref. (51), which may also depend on other parameters such as Hill coefficients, inhibitor constants, and stoichiometric coefficients, so these may also be necessary to determine the set of kinetic parameters $\mathbb{K}$, and by consequence, the model uniquely. With known rate laws, a model is also uniquely determined by the corresponding triple $(\mathbf{M}, \mathbb{K}, \rho)$.

Mass balance and the stoichiometric matrix S. For a stoichiometric matrix $\mathbf{S}$ including all reactions and reactants (internal and external ones), and the corresponding vector $\boldsymbol{w}$ of molecular masses (also known as molecular weights) of reactants, mass conservation within reactions implies

$$\boldsymbol{w}^\top \mathbf{S} = \mathbf{0}^\top \quad. \tag{76}$$

Note that, therefore, the vector of molecular masses must be in the left-null space of $\mathbf{S}$.

If we restrict the stoichiometric matrix to internal reactants i, then the internal product of $\boldsymbol{w}^\top$ with the columns of this new matrix with entries S_j^i is nonzero only for transporters

$$\begin{aligned} w_i S_s^i &\neq 0 \\ w_i S_e^i &= 0 \\ w_i S_r^i &= 0 \quad. \end{aligned} \tag{77}$$

Given that $\mathbf{M}$ is a mass-scaled version of $\mathbf{S}$, these relations are equivalent to Eqs. 2 in the main text; only transporters are capable of increasing or decreasing mass within the model.

We note that our considerations about mass conservation presuppose that all reactants also appear explicitly in the model (i.e. they have a corresponding row in $\mathbf{S}$). If some reactants (e.g. water or protons) are omitted from the model for convenience, mass balance is not satisfied. In fact, many models in the literature do ignore some reactants, in particular water; this needs attention in realistic models where water is not only a medium but also a reactant in many biochemical reactions.

Examples of GBA models and R code for numerical optimization. We present here 4 simple examples of GBA models. We assume for all models a simple irreversible Michaelis-Menten kinetics (72), so in each model turnover times τ are uniquely defined by a matrix $\mathbf{K}$ of Michaelis constants and a vector $\mathbf{k}_{\text{cat}}$ of turnover numbers for each reaction. Each row of $\mathbf{K}$ corresponds to a reactant, and each column to a reaction, matching the order in the matrix $\mathbf{M}$ (the entries for external reactants are separated by a horizontal line). See the corresponding ".ods" files for a more detailed description of the models, including labels for reactions and reactants and different growth conditions (i.e. external concentrations) for testing the model. Numerical simulations are done by running the R code in the file "GBA.R", with the variable "modelname" set to the desired model (e.g. modelname $\leftarrow$ "A"). Here $\mathbf{K}$ and ρ are in units of g/L, and k_{cat} in units of h^{-1} (resulting from product mass per protein mass per h). The code exports the results as a corresponding pdf file with relevant plots for visualization, and a csv file with the numerical values for the optimal cell state at different growth conditions.

Figure (5) presents the schematics and the corresponding parameters $(\mathbf{M}, \mathbf{K}, \mathbf{k}_{\text{cat}}, \rho)$ of each model. Metabolites are indicated with circles, and total protein with a rounded square. The numbers labeling reactants and reactions match the corresponding order of rows and columns in $\mathbf{M}$, with the last row corresponding to total protein and last column the ribosome reaction by default. All parameters are arbitrary, with the exception of the ribosome reaction where we use $k_{\text{cat}}^{\text{r}} = 4.55$ and $K_m^{\text{r}} = 8.3$ for its main substrate, based on the estimations for *E. coli* in Ref.(4), and $\rho = 340$ based on the measured *E. coli* dry mass density (52).

Models A and B have the simplest model structures (a linear pathway) for two and three reactions, respectively. Model C has a second transport reaction excreting metabolite "2". Model "D" has two redundant reactions ("3" and "4"), of which only one is active at optimal growth (see optimization results).
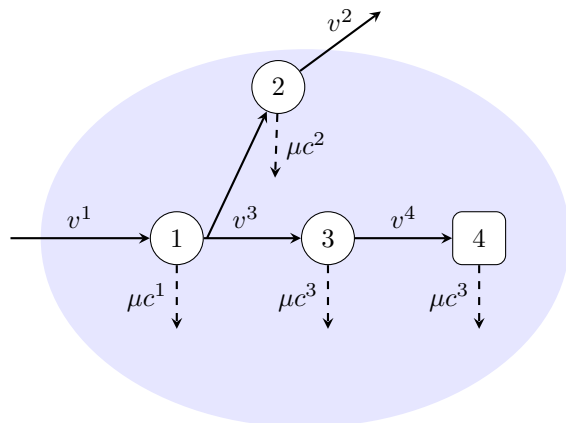
A
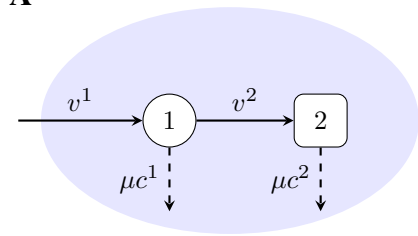
$$\mathbf{M} = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}, \mathbf{K} = \begin{bmatrix} 1 & 0 \\ 0 & 8.3 \\ 0 & 0 \end{bmatrix}, \mathbf{k}_{\mathrm{cat}} = \begin{bmatrix} 10 \\ 4.55 \end{bmatrix}, \rho = 340$$
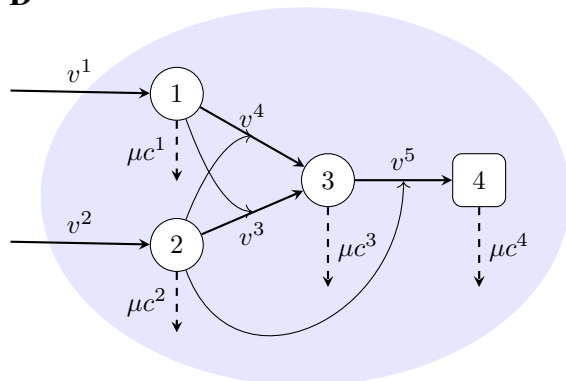
B
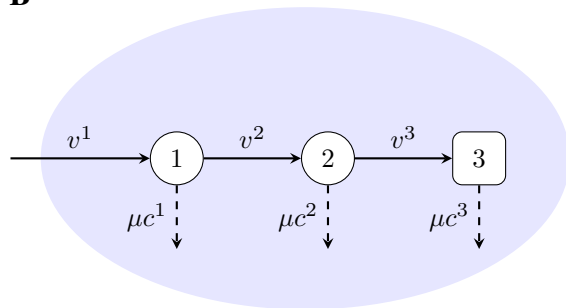
$$\mathbf{M} = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix}, \mathbf{K} = \begin{bmatrix} 10 & 0 & 0 \\ 0 & 10 & 0 \\ 0 & 0 & 8.3 \\ 0 & 0 & 0 \end{bmatrix}, \mathbf{k}_{\mathrm{cat}} = \begin{bmatrix} 10 \\ 8 \\ 4.55 \end{bmatrix}, \rho = 340$$

C

$$\mathbf{M} = \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & -1 & 0.2 & 0 \\ 0 & 0 & 0.8 & -1 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \mathbf{K} = \begin{bmatrix} 10 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 10 & 0 \\ 0 & 10 & 0 & 0 \\ 0 & 0 & 0 & 8.3 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \mathbf{k}_{\mathrm{cat}} = \begin{bmatrix} 10 \\ 100 \\ 8 \\ 4.55 \end{bmatrix}, \rho = 340$$

D

$$\mathbf{M} = \begin{bmatrix} 1 & 0 & -0.7 & -0.8 & 0 \\ 0 & 1 & -0.3 & -0.2 & -0.2 \\ 0 & 0 & 1 & 1 & -0.8 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \mathbf{K} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 10 & 5 & 0 \\ 0 & 0 & 10 & 5 & 10 \\ 0 & 0 & 0 & 0 & 8.3 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \mathbf{k}_{\mathrm{cat}} = \begin{bmatrix} 15 \\ 10 \\ 10 \\ 9 \\ 4.55 \end{bmatrix},$$

$$\rho = 340$$

Fig. 5. Schematics and parameters defining each model. For more details see the corresponding ods files in Dataset S1.

The dependence of λ_ρ on transporters. Equation 29 involves the sums $\gamma_j = \sum_i M_j^i$ of each column j in $\mathbf{M}$, which differ from zero only for transporters (Eq. 2). This means Eq. 29 can be separated into distinct equations for $j = \mathrm{r}, e, s$

$$(\partial_\mathrm{r}\mu)\,f_\mathrm{r} = 0 \quad , \tag{78}$$

$$(\partial_e\mu)\,f_e = 0 \quad , \tag{79}$$

$$(\partial_s\mu)\,f_s + \lambda_\rho \gamma_s\,f_s = 0 \quad . \tag{80}$$

From equations (78,79), we see that, at optimality, the summation in Eq. 30 can be simplified to a summation over s only

$$\lambda_\rho = -(\partial_s\mu)\,f^s \quad . \tag{81}$$

Substituting now the partial derivative given by Eq. 31, we obtain

$$\lambda_\rho = -\frac{\mu}{b^\mathrm{a}}\left(M_s^\mathrm{a}\,f^s - \mu\,\tau_s\,f^s - \mu\,f_l\,E_s^l\,f^s\right) \quad . \tag{82}$$

The first summand in the parenthesis equals to zero, since transporters do not produce protein ($M_s^\mathrm{a} = 0$), and the second summand can be expressed in terms of protein fractions $\phi^s = p^s/c^\mathrm{a} = \mu\,\tau^s\,f^s/b^\mathrm{a}$, resulting in

$$\lambda_\rho = \mu\left(\sum_s \phi^s + \frac{\mu}{b^\mathrm{a}}\,f_l\,E_s^l\,f^s\right) \quad . \tag{83}$$

Optimal enzyme concentrations and control coefficients. Equation 50 shares a striking analogy with an optimality condition for metabolic systems, established in (36): in an optimal metabolic state, maximizing a pathway flux at a limited total enzyme amount, the enzyme levels must be proportional to the flux control coefficients. This previous result reflects the assumption that the cell trades a cost (the sum of enzyme levels) against a benefit (the pathway flux), and that in an optimal state, the marginal cost and benefit, for any small change of enzyme levels, must be the same. Equation 50, for optimal growth states, has a similar interpretation, but without an explicit benefit function for fluxes. Here, instead, if an enzyme level in reaction A has an indirect effect on a flux in reaction B and makes reaction B proceed more efficiently, then catalyzing enzyme for reaction B can be saved (and resources be redistributed to increase growth). Hence, we now have a trade-off between a cost (of investing enzyme in A) and an "opportunity benefit" (enzyme saved in reaction B). In the equation, an enzyme of interest (in reaction A) is described by p_e, its effect on all fluxes in the network (reactions B) is described by the control coefficients C_e^l, and the catalysts of these reactions are represented by p_l. By summing all marginal "opportunity benefits" $p_l C_e^l$, we obtain the equivalent to the marginal flux benefit in (36).

From the similar Eq. 44 in the main text, we can infer that at optimality, no two active enzymes can realistically catalyze the exact same reaction (i.e., have identical columns in $\mathbf{M}$). If this were the case, then both turnover times τ_e would have to be exactly the same, since the marginal opportunity would be identical for both enzymes (see Eq. 24). This condition is highly unlikely to hold in realistic models, since real enzymes will always have different physical properties and therefore different kinetics. Thus, if several isoenzymes catalyze a particular reaction in a model, only one of these reactions will be active at optimality. The previous argument can be generalized to any two linear combinations of reactions in the model, and as a consequence, the submatrix resulting from restricting $\mathbf{M}$ to active reactions must have full column rank at optimal growth (see Refs. (4, 21, 22)). Note that in reality enzymes with very similar marginal values may still be expressed together due to the little selection pressure towards one of them. See model "D" on SI text "Examples of GBA models and R code for numerical optimization" for an example of a model with redundant reactions.

Amino acid composition of the *E. coli* proteome. Table S1 presents the amino acid frequency in the *E. coli* proteome, calculated from protein sequence (retrieved from genome annotation of *E. coli* NC_000913.3 from RefSeq (53)) and weighted by protein abundance measured in Schmidt et al. (47) under various growth conditions.

Table S1. Amino acid frequency in the *E. coli* proteome at various growth conditions.

Growth condition	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
Glucose	0.102	0.007	0.059	0.072	0.034	0.084	0.020	0.060	0.064	0.085	0.026	0.039	0.040	0.038	0.052	0.048	0.056	0.081	0.008	0.026
LB	0.102	0.007	0.058	0.074	0.033	0.084	0.019	0.060	0.068	0.081	0.026	0.038	0.039	0.036	0.055	0.046	0.056	0.083	0.007	0.025
Glycerol + AA	0.102	0.007	0.059	0.073	0.033	0.085	0.019	0.060	0.066	0.082	0.026	0.038	0.039	0.037	0.053	0.047	0.057	0.082	0.008	0.026
Acetate	0.105	0.009	0.059	0.069	0.035	0.084	0.019	0.059	0.061	0.085	0.025	0.039	0.041	0.041	0.047	0.049	0.057	0.078	0.009	0.028
Fumarate	0.104	0.008	0.059	0.070	0.034	0.084	0.019	0.059	0.063	0.085	0.026	0.039	0.040	0.039	0.049	0.049	0.056	0.079	0.009	0.027
Glucosamine	0.104	0.008	0.059	0.070	0.034	0.084	0.019	0.060	0.064	0.084	0.026	0.040	0.040	0.039	0.050	0.049	0.056	0.079	0.009	0.027
Glycerol	0.103	0.008	0.059	0.071	0.034	0.084	0.019	0.060	0.064	0.085	0.026	0.039	0.040	0.039	0.051	0.049	0.056	0.080	0.009	0.027
Pyruvate	0.103	0.008	0.059	0.071	0.034	0.085	0.019	0.060	0.063	0.086	0.026	0.039	0.041	0.038	0.049	0.049	0.056	0.080	0.008	0.027
Chemostat μ=0.5 h^{-1}	0.103	0.008	0.059	0.070	0.034	0.084	0.020	0.060	0.063	0.085	0.026	0.039	0.040	0.039	0.050	0.049	0.056	0.079	0.009	0.027
Chemostat μ=0.35 h^{-1}	0.105	0.008	0.059	0.069	0.035	0.084	0.019	0.059	0.062	0.085	0.026	0.039	0.041	0.040	0.048	0.049	0.056	0.078	0.010	0.028
Chemostat μ=0.20 h^{-1}	0.105	0.008	0.059	0.068	0.035	0.083	0.019	0.059	0.062	0.085	0.026	0.039	0.041	0.041	0.047	0.050	0.056	0.077	0.010	0.029
Chemostat μ=0.12 h^{-1}	0.106	0.008	0.059	0.068	0.035	0.083	0.019	0.058	0.061	0.085	0.026	0.039	0.041	0.041	0.047	0.050	0.057	0.077	0.010	0.029
Stationary phase 1 day	0.106	0.007	0.062	0.068	0.032	0.082	0.018	0.058	0.068	0.084	0.025	0.042	0.038	0.040	0.047	0.051	0.058	0.080	0.008	0.026
Stationary phase 3 days	0.105	0.007	0.062	0.068	0.032	0.083	0.018	0.059	0.067	0.084	0.025	0.041	0.038	0.040	0.047	0.050	0.058	0.080	0.008	0.026
Osmotic-stress glucose	0.103	0.008	0.060	0.071	0.033	0.083	0.019	0.059	0.065	0.084	0.025	0.040	0.040	0.038	0.050	0.049	0.057	0.080	0.009	0.027
42 °C glucose	0.102	0.007	0.060	0.073	0.033	0.084	0.019	0.060	0.065	0.084	0.026	0.039	0.039	0.038	0.052	0.048	0.057	0.080	0.008	0.026
pH6 glucose	0.103	0.008	0.060	0.070	0.033	0.083	0.019	0.059	0.065	0.084	0.025	0.040	0.039	0.038	0.050	0.049	0.057	0.080	0.009	0.026
Xylose	0.102	0.008	0.059	0.072	0.034	0.084	0.020	0.060	0.065	0.085	0.026	0.039	0.040	0.038	0.051	0.048	0.056	0.080	0.008	0.027
Mannose	0.104	0.008	0.059	0.071	0.034	0.084	0.019	0.059	0.064	0.085	0.026	0.039	0.040	0.039	0.049	0.049	0.056	0.079	0.009	0.027
Galactose	0.106	0.008	0.060	0.069	0.035	0.083	0.019	0.059	0.063	0.085	0.025	0.040	0.040	0.041	0.047	0.050	0.056	0.078	0.010	0.028
Succinate	0.104	0.008	0.059	0.070	0.034	0.084	0.019	0.060	0.064	0.085	0.026	0.039	0.040	0.039	0.050	0.049	0.056	0.079	0.009	0.027
Fructose	0.103	0.008	0.059	0.071	0.034	0.084	0.020	0.060	0.065	0.085	0.026	0.039	0.040	0.037	0.051	0.048	0.056	0.080	0.008	0.026
Coefficient of Variation:	1.25%	5.13%	1.48%	2.27%	2.44%	0.69%	2.61%	0.94%	2.87%	1.19%	1.01%	2.00%	1.82%	3.33%	4.15%	1.93%	1.09%	1.86%	7.55%	3.50%

SI_files.zip. Zip file containing files for numerical optimization, as described in the SI text "Examples of GBA models and R code for numerical optimization".